R E P O R T R E S U M E S

ED 017 793 AC 001 541

PSYCHOLOGICAL RESEARCH IN CLASSROOM LEARNING, REPORT ON A CONFERENCE (BROMWOODS RESIDENTIAL CENTER, WASHINGTON UNIVERSITY, ST. LOUIS, MO.).

BY- WIENTGE, KING M., ED. AND OTHERS
WASHINGTON UNIV., ST. LOUIS, MO.,SCH.OF CONT.EDUC.
REPORT NUMBER R-PUB-11 PUB DATE 67
CONTRACT OEC-3-7-061134-1532
EDRS PRICE MF-$0.50 HC-$5.04 124P.

DESCRIPTORS- *ADULT LEARNING, *EDUCATIONAL PSYCHOLOGY, *CLASSROOM ENVIRONMENT, *INSTRUCTIONAL TECHNOLOGY, *EVALUATION, RESEARCH, HIGHER EDUCATION, MILITARY TRAINING, FLIGHT TRAINING, AEROSPACE TECHNOLOGY, CLASSROOM TECHNIQUES, CLASSROOM PARTICIPATION, SIMULATION, PACING, SYSTEMS APPROACH, TESTING, MEASUREMENT INSTRUMENTS, UNITS OF STUDY (SUBJECT FIELDS), RETENTION, STATISTICAL DATA, PROGRAM ATTITUDES, PROGRAMED INSTRUCTION, COMPUTER ASSISTED INSTRUCTION,

PAPERS WERE PRESENTED AT A CONFERENCE ON CLASSROOM LEARNING ON SUCH TOPICS AS PROGRAM DESIGN, TESTING, AND OTHER EVALUATION TECHNIQUES, COMPUTER ASSISTED INSTRUCTION, PROGRAMED INSTRUCTION, SIMULATION, PACING, AND RETENTION. SEVERAL TREATED MILITARY TRAINING, ADULT LEARNING, AND ADULT-CENTERED CLASSROOM TECHNIQUES. IN ONE PAPER, THE SYSTEMS APPROACH TO ADULT LEARNING WAS PREDICATED ON THE APPLICATION OF ENGINEERING LOGIC IN FORMULATING OBJECTIVES, IDENTIFYING AND ALLOCATING FUNCTIONS, AND PLANNING FOR OPERATION OF THE SYSTEM AND SYSTEM TESTING. ANOTHER STUDY COMPARED THE PERFORMANCE OF NAVY TRAINEES IN RADAR AND RADAR NAVIGATION TECHNOLOGY AS MEASURED AT FOUR POINTS IN THE TRAINING SEQUENCE. ONE RETENTION STUDY COMPARED PROGRAMED AND CONVENTIONAL INSTRUCTION IN A NAVY TECHNICAL COURSE. ATTITUDES OF INSTRUCTORS, TRAINING ADMINISTRATORS, AND NAVY AND MARINE ENLISTED MEN TOWARD PROGRAMED INSTRUCTION WERE SURVEYED AT EIGHT AVIATION TECHNICAL TRAINING SCHOOLS. A STUDY OF ELECTIVE PARTICIPATION IS CURRENTLY BEING MADE. THE DOCUMENT INCLUDES TABLES, DIAGRAMS, AND A BACKGROUND PAPER. THESE PAPERS WERE PRESENTED AT A CONFERENCE HELD AT THE BROMWOOD RESIDENTIAL CENTER, WASHINGTON UNIVERSITY, ST. LOUIS, MISSOURI. (LY)

AC001541
Apr 68
ED017793

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

PSYCHOLOGICAL RESEARCH in CLASSROOM LEARNING

Edited by
King M. Wientge
Philip H. DuBois
and
Harry Gaffney

SCHOOL of CONTINUING EDUCATION
RESEARCH PUBLICATION
WASHINGTON UNIVERSITY
ST. LOUIS, MISSOURI

AC001541

SCHOOL of CONTINUING EDUCATION RESEARCH PUBLICATION
WASHINGTON UNIVERSITY
OFFICE of RESEARCH,
SCHOOL of CONTINUING EDUCATION
and the DEPARTMENT of PSYCHOLOGY
ST. LOUIS, MISSOURI 63130

PSYCHOLOGICAL RESEARCH IN CLASSROOM LEARNING

A Report on a Conference Held at the
Bromwoods Residential Center
of Washington University
Sponsored by the Department of Psychology and the
Office of Research, School of Continuing Education,
Under Office of Naval Research Contract Nonr 816(14) and
U. S. Office of Education Contract No. OEC 3-7-061134-1532,
Research Project 6-1134

Edited by:

King M. Wientge

Philip H. DuBois

Harry Gaffney

School of Continuing Education Research Publication

Number 11

1967

FOREWORD

Research interest in the classroom, focus of formal education from the kindergarten to the graduate school, appears to be increasing. Long-accepted classroom methods are being challenged by new concepts, procedures, and devices.

It is appropriate that innovations be tested and compared with traditional methods. To this end Washington University has been conducting research at Naval Air Station, Memphis, under Contract Nonr 816(14) and on its own campus through the Research Office of the School of Continuing Education, under Office of Education Basic and Applied Research Project 6-1134. Personnel directly interested in these projects and several outside consultants met for a two-day conference at Bromwoods Residential Center. This document reports the formal presentations and subsequent discussions.

It seemed appropriate to invite an anthropologist to establish the general framework for a look at modern education. The initial paper provided a perspective on methods used in primitive societies in insuring youth acceptance as a member of society.

Another broad view was provided by a paper on a systems approach to education. After discussions of the measurement of learning, attention was given to computer-assisted instruction and to the role of simulation.

Other topics covered in the conference were the new look in educational techniques in the public school, in technical training and in colleges and universities, and reports on seven research studies which were in progress or recently completed.

Acknowledgment is made of assistance under the research contracts mentioned above and to the personnel of the School of Continuing Education who made the arrangements for the successful meeting at Bromwoods: Dean Lynn W. Eley, Dean John B. Ervin, and Mr. Andrew Thomas.

King M. Wientge
Philip H. DuBois
Harry Gaffney

TABLE OF CONTENTS

Page

VISIT TO CLASSROOM ISLAND: AN ANTHROPOLOGICAL REPORT
Charles E. Thomas 1

A SYSTEMS APPROACH TO ADULT LEARNING
Robert M. Gagné 6
Discussant: Earl I. Jones 20

MEASUREMENT OF THE PRODUCTS OF TRAINING
Philip H. DuBois. 22

STRATEGIES OF ACHIEVEMENT TESTING
Winton H. Manning 28
Discussant: J. Roger Berkshire. 36

COMPUTER ASSISTED INSTRUCTION: SOME FACTS AND FANCIES
James J. Regan. 38
Discussant: Robert M. Gagné 50

SIMULATION APPLIED TO EDUCATION
Edward R. Jones 53
Discussant: James M. Vanderplas 67

THE NEW LOOK IN EDUCATIONAL TECHNOLOGY
IN THE PUBLIC SCHOOLS
James M. Dunlap 70
IN TECHNICAL TRAINING
G. Douglas Mayo 72
IN COLLEGES AND UNIVERSITIES
Marion E. Bunch 75

PROJECT ELECTIVE PARTICIPATION: A PRELIMINARY REPORT
King M. Wientge 81

LEARNING OPERATIONAL EQUIPMENT AS A CRITERION IN TRAINING RESEARCH
G. Douglas Mayo 86

EVALUATION OF A PARTIALLY SELF-PACED COURSE
Kirk A. Johnson 94

THE QUIZ AS A TEACHING DEVICE
Leon Clodfelder 99

A RESIDUAL GAIN MEASURE OF REMINISCENCE
Edward N. Peters. 106

A STUDY OF RETENTION OF PROGRAMMED LEARNING
William C. Orr. 109

CONSUMER ACCEPTANCE OF PROGRAMMED INSTRUCTION
Howard L. Fleischman. 112

CONFERENCE PARTICIPANTS

J. Roger Berkshire, Naval Aerospace Medical Center, Pensacola

Clay V. Brittain, U. S. Office of Education

Marion E. Bunch, Washington University

Leon Clodfelder, Washington University (stationed NAS, Memphis)

Philip B. DeVries, Jr., Washington University

Philip H. DuBois, Washington University

James M. Dunlap, University City Schools

John B. Ervin, Washington University

Howard L. Fleischman, Naval Air Station, Memphis

Harry Gaffney, Washington University

Robert M. Gagné, University of California, Berkeley

Mary Grohmann, Washington University

Earl I. Jones, Navy Training Research Laboratory, San Diego

Edward R. Jones, The McDonnell Douglas Company

Kirk A. Johnson, Naval Air Station, Memphis

Robert E. Lefton, Psychological Associates, Inc.

Winton H. Manning, College Entrance Examination Board

G. Douglas Mayo, Naval Air Station, Memphis

William C. Orr, Washington University

Edward N. Peters, Washington University

Joseph G. Pfeffer, Washington University

James J. Regan, Naval Training Device Center, Orlando

Robert N. Schneider, Jefferson Barracks

Donald R. Schuster, Washington University

Myron A. Spohrer, Washington University

Andrew E. Thomas, Washington University

Charles E. Thomas, Washington University

James M. Vanderplas, Washington University

King M. Wientge, Washington University

VISIT TO CLASSROOM ISLAND: AN ANTHROPOLOGICAL REPORT

Charles E. Thomas

Anthropology is a "live-in" science. An anthropologist typically moves in with his subjects in a primitive society, but now that we are running out of tribes, becoming technologically unemployed as it were, we are moving into the field of urban anthropology, which means that we are moving in on the sociologists, which in turn means that we are taking on all the paraphernalia of the sociologists. Nevertheless, the tradition has been that if a person wanted to join the academic foreign legion, he became an anthropologist and left his wife a long time. Out of this grew a tremendous amount of empirical data—a tremendous respect for the richness and exotic qualities of empirical data derived from living-in with societies.

I would like to describe how several primitive societies met the challenge of education, primarily at adolescence when the young person became a full fledged adult and had to assume the responsibilities and expertise of his society. How did a variety of societies meet this large and critical educational challenge?

The first example that I would like to deal with is the Copper Eskimo inhabiting the Canadian tundra below the Arctic Circle. It is a society that lives in a difficult environment, and while it has rather less complex standards than our own, it does have a rather advanced and ingenious technology—a great deal has to be learned in the art of herding reindeer and the art of hunting seal, both vitally important in sustaining life. The population is inclined to die off very easily. In this frail but at the same time technologically complex society, the Eskimo treats the learning process as supportive-integrative. From his very earliest cognizance of what is going on around him, the child is involved in learning. When the child is two the Mother throws him a few pieces of thong and teaches him, much as in Playskool games, to thread the thong through the boot. On a 15-mile walk in the snow the thong has to be threaded a certain way or the boot will come loose and then there is trouble. When uncles and other relatives come to visit at the house, conversation has to do with learning. For example, "How is Abu getting along with harpoon throwing?" From earliest childhood up to maturation and marriage there is continuous involvement in learning, learning in the family context with all the overtones of belonging and acceptance.

Learning is really undifferentiated from everyday aspects of life, that is, work and play. Accordingly, the family represents the nucleus of education.

Now, consider another society that provides contrast—the Pilaga of Brazil. The Pilaga are of the opinion that not the older members of the family, but the peer group is the primary source of education. This is the progressive school approach to education, and has the same idea as the student university. One can't dig the old folks because they are not really with it. One can most easily learn from interaction with one's own siblings and peer groups. It is interesting, too, that among the Pilaga the primary means of learning the important roles comes from a distinct and clearly observed role playing. They learn to hunt by going out in childrens' hunting parties just like their fathers and the older men, but they hunt rats with a small bow and arrow instead of going out to hunt wild pig. As a boy or girl they learn about agricultural practices by having their own little field to cultivate. They learn about the social mores of the society by playing house—literally acting out all of the various activities of that society. It is interesting that Pilaga children play house, simulating sexual intercourse—Jules Henry was the one who did much of the research corroborating the manifest sexuality of children. It seems that sex is one of the primary reinforcement techniques employed by the Pilaga in age-set sibling education.

If this represents the sibling form of education and if the Copper Eskimos represent an integrated family approach, most of the tribes of the South Seas represent a form of education much closer to our own. The great men, the graybeards, the high priests, and the elders sit in council. Teaching in one's peer group or in one's family is discredited and it is really only the full professors and associate professors who are allowed high status. I think the Polynesians could move right into the center of things on an American campus!

In primitive society the most dramatic and most oft recurring type of education is what I would call the sink-or-swim traumatic type. Negative reinforcement is part and parcel of the creation of the mature adult. This is typified by the Plains Indians, a warlike peoples who rear children in a very segregated situation. The child is breast fed, picked up whenever he cries, played with, constantly spoiled by older brothers and sisters and loved continuously by the mother until he reaches puberty. At that moment he is taken out of the female situation, moved over to the companionship of men with whom he has never had much contact. He is taken up on a mountainside and about his only education to that point is the absolute knowledge that evil spirits reside in those mountains. He is given a small supply of water in a jar, and he must spend three days, in what the equivalent for us would be a graveyard vigil. He must come down in the cold with nothing but a small serape to wear over his shoulder. He must return at the end of the third day to participate in a ritual in which he sees gods with great frightening masks dance in front of the fire. He has been told that these gods came from remote parts of the mountains, that they are in fact divinities. Then he sees the patriarchs of the village take off their masks and say it has been a sham—that they have been impersonating the gods. On top of this he is then subject to being cut with various stone implements, and finally to the passage of thorns through his ear lobes, his mouth, other parts of his body, and frequently circumcision. Within another year he will go through probably the most grueling rite in this intense period of education, which is the Plains Indians' Sun Dance. A central pole is erected with long ropes which are attached to the young warrior by skewers through his muscles and skin. He then dances around the pole pulling at its ropes, finally tearing these through his muscles as final proof of his manliness. This traumatic form of learning is an integral part of this no-nonsense society. It is a society in which the warrior must be tested, and tested successfully, long before he goes into battle. The society also provides that those who default and are not able to go through the long rituals return to the camp of the women, put on woman's garbs and then pick up all of the roles in that particular society that neither men or women perform. Women are not allowed into the men's house and men are not allowed to do housekeeping work so the transvestites do the work.

This relates to what is common in primitive society—an overwhelming experience that comes in a short period of time as a traumatic sink-or-swim form of education. In the last several months I have been talking to students about the master teacher on our campus. I have been inclined to feel that there is a close parallel between this type of educative experience and one of the most prevalent master teacher groups. This is the great man—the man who traumatizes students from an aloof position, spelling out the rules. He has a compendium-like book list. He demands erudite term papers in introductory courses and his final examinations are exercises in the total discipline in which he is involved. This man frequently surrounds himself with a cordon of graduate students. His office hours are frequently very short and then unattended, so that one gets "Won't you speak to Mr. So and So today, the professor is not here." Here is a stern remoteness, a combination of an Olympian deity and a castrating father. Students may hate such a person in class only to find in retrospect that this was their greatest hour while they were on campus. Also, students often say that a teacher must be outstanding to play a highly punitive and arrogant role. If he does not have genius and inspiration, it falls flat and becomes a sham.

A last type, which seems to represent a combination of elements, I would like to talk about is the Arunta of Australia who inhabit an extremely difficult environment. This is a stone age hunting tribe that lives in a semi-arid region. It has a very high mortality rate. In this group, much like the Plains Indians, there is an attempt to create a coddled environment up to the time of the puberty rituals. Among the Plains Indians, however, one had to go it alone. The teachers themselves were remote and isolated from the trauma of the young neophyte who is going through these puberty rituals. In the case of the Arunta, there is a highly supportive role played through the most traumatic forms of ritual by the father and the uncles of the child from whom, again, he has been isolated.

The ritual is something like this. The child, after spending most of his childhood with the women, is finally taken by the father and the uncle to the cermonial area. He is painted elaborately in the traditional and secret symbols of the society. He is then put on a blanket and tossed in the air— a really frightening experience. Then he is involved in a painful series of trials by fire with searing of the flesh, bloodletting, passing of thorns through soft parts of the body, and circumcision. Finally, the ceremonies are completed with the relativesbiting the young boy on the head so that the blood flows. This type of ritual seems a case of cognitive dissonance in that the father exhibits tears and protestations, at the same time maintaining that the dangerous ritual is necessary.

On a campus between the two roles of the supportive-integrative, and the traumatic sink-or-swim professor, one finds the type exhibiting cognitive-dissonance, a type that attempts to adhere to rigid standards but also to play a supportive role. I am impressed that it is not this particular type of faculty member that is seen as most successful, but rather the traumatic sink-or-swim great man who is the one most frequently admired. I would like to conclude with the statement that in primitive societies there is a high correlation between the severity of the environment and the need for successful education at an early age. The severity of the environment creates concomitant severity in the rites of puberty and of educational techniques associated with them. It would seem to me also that the primary characteristic in making some parallel between primitive societies on the one hand and great teachers on the other is that great teachers invoke and stimulate tremendous emotional involvement. Love, hatred, anxiety, hostility, belonging to family or tribe—all of these basal forces seem to be evoked by successful teachers. At this point I feel, as an anthropologist, some concern for the modern mechanistic techniques of education, which in and of themselves may be creating more efficient means of learning, but because they are mechanical rather than human, they are providing frail, thin, and ineffectual forms of reinforcement. At the same time they deny the powerful forces which the anthropologist, cross-culturally and even within our own society, sees as the primary motivating, involving force in the whole educative process.

It is with this anxiety that I came to you and it is with the same one that I take my leave.

A SYSTEMS APPROACH TO ADULT LEARNING

Robert M. Gagné

Can one apply the methods of systems design and development to adult learning? Yes, indeed one can. In fact, the current situation and urgent requirement for a comprehensive course of action in this field makes adult learning a "natural" for the application of systems design methods.

Let me be clear, first of all, that I am talking about systems design and development, and not about systems analysis, which is a much more specialized technique. One can, it is true, make an analysis of the current situation pertaining to the learning and continuing education of adults, using techniques of operations analysis or the even more specialized techniques of cost effectiveness. But although the results of such studies might well contribute some valuable information, it would be my contention that the considerably broader point of view of systems design is needed. In simple terms, we can see in broad outline what is wrong, and what is missing. What we need is a thorough and systematic effort to design a system which meets the needs of adults far better than what we now have. In other words, we need to put together all of the technology we can identify, use all of the theoretical predictions about adult learning we can make, and attempt to design a new system which will approach optimal effectiveness.

The design of a system begins with a statement of purpose. The system is designed to accomplish something, a goal which is agreed upon by the society within which the system is to work, or by some segment of it. Such a purpose, or perhaps a set of purposes, is rationally analyzed into certain subordinate goals which cover the various forms of operation the system is to have. (If there is only a single form of operation, well and good, but there are more likely to be several.) At about this point, one begins to talk about the functions of certain parts of the system, which may be called components, or if they are larger entities, subsystems. Also, at about this point, one has to introduce concepts like input and output, to refer to the actions of these components. A system is organized, after all, by planning and bringing about some sort of compatibility, or "match" between the output of one component and the input of the next in a sequence of operations. In a rocket system, for example, a match must be made between some planned output of a "first stage" of propulsion and an input to the "second stage", in order that the latter stage can be put into operation. Or, to take a human example, a "match" must be designed between the numbers of people using a mass transit system at any given time, and the capacity of the vehicles which transport these people. Input-output considerations of this sort not only clarify what is meant by the function of any component, but also lead to decisions generally called allocation of functions to different kinds of components.

I mention these things mainly to emphasize that the design of systems is subject to engineering logic. This is not the same as scientific logic, for a scientist is not ordinarily concerned with purposes and functions. But these are the kinds of decisions an engineer is supposed to make, and is used to making. Nor do I mean to imply that because engineering logic is used, one must be concerned primarily with hardware. There are many systems and subsystems which are composed primarily of people, rather than hardware—for example, the system of student counseling in schools, or the system of trading in stocks in a "market." It is engineering logic which has designed such systems—a logic which deals directly with goals, purposes, objectives, functions, inputs and outputs. One can, of course, substitute another language if this one seems to specialized. But the logic will, I believe, be quite the same.

Can one, then, apply such engineering logic to the design of a system of adult learning? It will be my purpose here to show that this can be done, and to describe the implications this has for action. But before I launch directly upon that task, it may be well to review briefly where we now stand with respect to adult learning, and what are some of the reasons for believing we need a new system at all.

The Problem of Adult Learning

Learning occurs throughout life. The increasing complexity of the society in which we live and the accelerating rate of social and technical change are increasing the importance of learning later in life. The rate of technological and social change soon renders obsolete some of the skills and knowledge which have been achieved by the end of formal full-time education. But it is not easy to keep learning in all of the appropriate areas of later life. Gardner (1963, pp 8-9) in speaking of the commencement lecturer's adjuration to "keep on learning" says:

> "Unfortunately the commencement speakers never tell us why their advice to keep on learning is so hard to follow. The people interested in adult education have struggled heroically to increase the opportunities for self-development,

> and they have succeeded marvelously. Now they had better turn to the thing that is really blocking self-development--the individual's own intricately designed, self-constructed prison, or to put it another way, the individual's capacity for self-renewal."

Not only is it difficult to maintain a capacity for self-renewal; the increasingly complex roles of citizen and worker demand more than casual or incidental new knowledge. Organized systems of knowledge are achieving a place of importance for all elements of our society which has never before been experienced.

The importance of adult education to our society is likely to be enhanced in still another way by increasing trends in technology and automation. This is the need for the individual to build intellectual bridges between humanistic values, on the one hand, and the methods and accomplishments of technology on the other. Liveright (1964) sees the filling of such a gap as one of the major purposes of continuing education. Technologists, he believes, must be exposed to opportunities for humanistic studies and exploration, while humanists at the same time seek a more profound understanding of the values and goals of technology. For those people who consider themselves to have finished regular school attendance, the bringing together of the "two cultures" may have a profound effect on successful living in today's world.

Some Historical Points about Adult Education

Despite its increasing importance, the direction for adult education is not necessarily clear. Many types of adult education have been developed to meet a variety of perceived learning needs. They have been sponsored by universities, clubs, churches, unions, libraries, museums, private schools, government agencies, and public schools (Hendrickson, 1960). They have varied from occasional lectures to involved sequences of courses leading to diplomas; from intensely vocational orientation to pure learning-for-learning's sake; from highly organized curricula to loosely defined discussions; and from literacy training to the theory of relativity.

It is also true that programs of adult education have served various purposes at various times and places. On some occasions, for example, such programs were undertaken largely for "citizenship" purposes; this was particularly true back in the early part of this century, in those communities which received many immigrants. At other times, the emphasis has been largely "occupational," with the attempt to provide a means for the working man or white-collar worker to acquire additional skills in his job or profession, and thus to advance himself within that occupation. There are many offerings in adult education today which serve this purpose. Still a third goal for the education of adults, and one that has continued for many years, may be identified as "self-fulfillment," which is of course only a small part of what Gardner means by self-renewal. Many offerings in this field are concerned with arts and crafts, including such courses as how to weave, design of pottery, as well as appreciation of music and art.

How adequately these courses have met a real need is, in general, unknown. The agencies which offer courses must satisfy their customers, the students, at least to the point of attracting enough to make continuation seem worthwhile. On the other hand, many agencies would reject a "customer is always right" philosophy, on the grounds that the educational institution should be striving to lead its students toward higher goals rather than merely responding to the goals which the students are able to express. In any case, choices must be made in developing the curriculum, and the institution must find ways of making these choices rationally rather than accidentally. No really systematic basis currently exists for selecting adult education objectives or for developing curricula once they are defined (Luke, et al, 1959; Sillars, 1958).

Superintendents of public schools in large majority subscribe to the idea that adult education should fulfill multiple purposes (National Education Association, 1964). Perhaps partly as a result of these broad aims, however, it is recognized that the curriculum in adult education often falls short of meeting the needs of adults. For example, a Commission of the Professors of Adult Education has said (1961):

> "To its great disadvantage, the curriculum of adult education has no plan. It is largely an à la carte menu comprised of miscellaneous items. It is not organized in a way that provides continuity and integration of learning. It is need-meeting but not goal-fulfilling."

The Systems Approach

It is quite evident that the application of a systems approach to adult learning must begin with a study to determine goals. This is necessary because it is from a statement of goals of the system that all the logic of this approach flows. One could, of course, have these goals determined by a panel of "experts," and this would not be an entirely unreasonable approach. But insofar as a program of adult learning is to be instituted by a community or other segment of our society, it seems a wise thing to do to attempt to determine what the needs of adults in this segment of society are.

However, it should be noted that I do not think such an effort can be done by means of the more superficial "survey" techniques. Rather a highly sophisticated line of questioning should be followed, in order to come as close as possible to "true needs." It seems to me that this simply cannot be done by asking questions like the following:

"What courses would you like to take in adult education?

_____ Pottery
_____ American Civilization
_____ Modern Poetry
_____ Stenography
_____ Woodworking"

In order to get closer to true needs, one would have to ask questions based upon the premise that learning means a change in capability. From this basic assumption, one can then proceed to formulate the kinds of questions that are designed to reveal concrete instances of desire (on the part of those questioned) for changing their capabilities. Only when one had collected a sufficient number of these concrete instances of felt needs would one be justified in forming them into rough descriptive categories. Such categories might be vastly different from the names of courses one now finds offered. In this process of determining the goals of adult learning, one should not impose pre-existing, and possibly quite obsolete, categories upon the results.

Another set of operations is also needed in determining goals. Such goals, particularly in today's world of rapid societal change, must be forward-looking. And results based upon adults' concrete experience cannot in themselves embody this kind of orientation to the future. One must, therefore, somehow take account of the changes in communications and media, in computerization of work, and in probable increases in leisure time, all of which we are constantly being reminded of by our present popular media. For this purpose, one needs some kinds of predictions of things to come and of the trends which get us from here to there. It is difficult to see how such trends are to be described without some wise "experts." It would be nice not to have to use them, because their reliability is suspect, but they seem to be necessary. We must, I think, depend upon the projections of broadly informed and intelligent people to give us estimates of how much non-occupational time we will have in the future, of how many of our tasks will be performed by computer systems, of how our daily lives will be affected by changes in the arts of communication.

I should judge that goals for adult learning could best be determined by means of a rational interaction of these two sources of information. From the past, we would take the most highly concrete and empirically based information we could find--derived from answers to questions about actual needs. From projections into the future, we would use the most highly informed wisdom available. But it would be a rational melding of these two kinds of information which would give us the best formulation of the broad goals of a system of adult learning.

Formulation of Objectives

Perhaps I have said enough about goals to indicate in broad outline what seems to me a sensible approach to their determination. At the least, it will be clear that I think this is an essential first step. If we are to design a system, we must know what its purposes are to be, what outcomes are to be expected of it, in order to proceed with system development.

The kind of study I have suggested would lead, then, to broad goals. But they would still be broad. They need then to be categorized further in terms of the human functioning, or the classes of human performance, which make them up. For example, one would need to develop categories of what kinds of human capabilities are involved, for the range of human adults included, in the broad goal of increased occupational competence. These might include such categories as communication capabilities, human relations capabilities, numerical and quantitative capabilities, and so on. Such performance classes might bear some relation to the categories of "academic" subject-matters, such as language and mathematics. But one should not, and need not, prejudge this issue. Just because these traditional categories have grown up around the education of youth, and seem therefore to represent a kind of stable reality, is no reason to suppose that exactly the same categories would emerge in consideration of adult learning requirements.

At any rate, broad goals would need to be broken down into more specific categories of objectives. Let me state that learning, if that is the purpose of the system, cannot be conceived as a process without relating it to observable performances. And further, if a system is to be designed, we must design it with the aim of producing certain specific outcomes, and that is simply another name for objectives.

System Design

Having defined system outcomes, one can more or less readily perceive how the usual steps in system design may be carried out. These are, in order, as follows:

1. Subsystem functions to be performed. One must identify the major parts of the system (i.e., the subsystems) which can be put together to accomplish the systems objectives. In all probability, these would include a subsystem of administration, one for instruction, one for guidance, and one for evaluation.

2. Allocation of functions within subsystems. One has to plan how each subsystem will work, what kinds of components it will contain, and how these functions will be allocated to the available components. In system language, outputs of components must be "matched," or made compatible, with the inputs of other components.

The instructional subsystem, for example, must involve consideration of how the following functions are to be performed: measuring entry capabilities of students; providing for differences in rates or styles of learning; determining proper sequences of learning; matching media with learning requirements; providing for retention and transfer of learning; assessing learning outcomes. Obviously, some of these functions can best be done by human beings, some by printed communication media, some by pictorial media, and some by various combinations of these. Similar remarks could be made concerning other subsystems involved in the total system for adult learning.

3. Identifying components. At first glance, it would appear that a wealth of components is available, and that the job is mainly one of selecting the right ones. After all, we have many kinds of instructional devices, including textbooks, motion pictures, television, projected pictures, and many others. We have many kinds of tests which can be used for student assessment and evaluation. And we have much categorized information pertaining to the guidance of the individual student.

As is often true in system development, however, closer examination of these components may show that many of them were designed for other purposes, and may require either extensive adaptation, or actual redesign, when considered for use of a system to promote adult learning. For example, to what extent are textbooks formed to meet certain literary "story-telling" standards, as contrasted with efficiency of communication, or efficiency of learning? To what extent have motion pictures and television programs been designed specifically with artistic criteria in mind, as opposed to effectiveness in instruction? Are slides designed as pretty pictures rather than as instructional tools? One might even ask, to what extent has the oral communication of the teacher been designed to accomplish a kind of entertainment task, as opposed to the task of specific behavior modification?

In designing a workable and effective system, questions like these must be faced squarely, and answered honestly. Components which will not accomplish the necessary functions must either be abandoned, or modified so that they will. This aspect of system development will not be accomplished well unless it is possible to ignore both sentimentality and tradition.

Plan for system operation. Having identified the functions of components, and selected or modified those which will perform these functions, it is now possible to put them together in such a way as to accomplish in a progressive fashion the goals intended for the system. Sometimes a flow-chart is used for this purpose. Let me instead attempt to describe system operation in a series of steps, as follows:

Step 1. Adult student appears at Adult Learning Center evincing interest. Records of vital statistics are made.

Step 2. Student is given an initial guidance interview, in which information about his interests, life goals, previous education, etc., is obtained, and he is given general information orienting him to the system.

Step 3. On the basis of information obtained in Step 2, student is given a set of tests designed to measure his level of achievement in certain broadly defined areas of human competence. For example, what mathematics can he do, what history can he recount or interpret, what tools can he use, what sort of writing can he perform, etc. (It may be noted that tests with these characteristics probably do not now exist, and would have to be developed).

Step 4. Student enters a second guidance interview, where the results of testing are available. What is to be achieved is a match between the student's goals and the kinds of instruction available.

Step 5. Student enters upon a program of instruction, let us suppose in three different areas. In each of these, the first thing done is measurement, this time to determine what may be called his "entering level" of competence in the specific area. On the basis of this test, the teacher (functioning as a manager of instruction) makes an initial assignment.

Step 6. Student begins his study of each area. Many resources are presumably available for such learning. However, these are not allocated at the student's will or preference. Rather, they have been carefully matched to the learning task to be accomplished, and to the conditions of learning required for this task (cf. Gagné, 1965a; Briggs, Campeau, Gagné and May, 1966). Thus, pictures are used when it is known that learning will be facilitated by their use in place of verbal descriptions (e.g., in showing the operation of a gasoline engine); whereas verbal presentations are used when they will accomplish a similar purpose (e.g., in representing the questions asked by a housewife to a television repairman).

Step 7. Student engages in a variety of learning situations in attaining successive levels of competence. All kinds of media are employed. At frequently placed "stages", assessment is undertaken to determine whether the student has attained the desired competencies. If he has, he is given a new assignment and sent on to the next stage in learning; if he has not, his specific deficiencies are identified, and he is encouraged to re-study.

Step 8. Periodically, the student is given an opportunity to join with other students in teams, seminars, or other groups for the purpose of developing new ideas or creating new products. The most important criterion for joining such groups is the prior attainment of suitable skills or knowledge relevant to the activity to be undertaken by the group. This is essential. Otherwise, the attainments to be aimed for by the group are not highly specified. They might decide, for example, to build an electric automobile, or to write a book, or to formulate a petition.

Step 9. Once or twice a year, the student is again assessed by suitable testing, in order to provide him with an indication and a record of his progress. Such assessment will again be followed by another guidance session, in which goals may be reviewed and revised if this is found desirable.

Returning now to the major set of stages in system design and development, there remains only one more to cover:

System testing (evaluation). This phase of system design should probably take as its model the engineering test, particularly the kind called a "field test." The question to be asked is: to what extent is the system meeting the objectives set for it? I believe we have the techniques to carry out such a test, although we don't often use them. Evaluation of the system probably needs to consider obtaining measures of at least the following: (a) kinds and amounts of changes in student competencies; (b) evidence of student motivation to engage in continued learning; (c) evidences of students' accomplishments, and particularly their increased goal-orientation and creative activity; and (d) evidences of realistic goal-setting on the part of students.

Now that I have given some description of the system as a whole, perhaps it will be well to review briefly some of its characteristics.

1. The system assumes adult motivation to enter into the system. Such motivation may of course be various, but they are all enlisted in the service of learning.

2. The central operation of the system is increasing student capabilities (or otherwise changing his mental dispositions); the central component of the system is the student.

3. The instructional subsystem is designed to meet the needs of the individual student, and for no other purpose. Learning is by and for the individual.

4. Incentives are so arranged that initially varied motivation is channelled into a desire for achievement, a "love of learning," and social approval for created products.

5. There is frequent assessment, of a sort and for purposes that are probably not met by currently existing instruments. The emphasis in measurement is on "competency status," or "what the student can do."

6. A subsystem of guidance, emphasizing self-confidence and self-dependence, performs integral and important functions.

7. Systen evaluation is planned from the beginning, and is also an integral portion of the system.

Some Critical Points

Now that the system as a whole has been briefly described, I can turn attention to certain critical issues in its design, and try to point out how they may be related to and illuminated by certain findings of research on the learning process.

How should a system of adult learning differ from a system of learning for children? Such a contrast may help to define the kinds of instructional objectives to be achieved by adult learning.

Basic Skills. A system of learning for the child surely must take into account the fact that certain basic skills have to be learned first. These skills, which include reading, writing, and figuring, along with certain others, are considered basic because they are essential to the process of continued learning itself. In contrast, it would be desirable if a system of adult learning were able to assume these basic skills. However, the latter system may have to have special "remedial loops" for some of them. If one is interested in the broad range of the adult population, the probability that some people cannot read painlessly, cannot add fractions, cannot speak standard English, and cannot compose a simple written communication, has to be faced. Designing such instruction takes a good deal of ingenuity. The materials should be designed to interest adults, mainly, I think, by representing problems which are as real to life as can be managed. An example would be the task, in written composition, of "describing to a novice how to fly a kite," as opposed to the task of explaining "what flowers mean to me."

Acquiring knowledge. The system of education for the child requires that he code, organize, and store a great deal of what is called "knowledge." These are facts and rules which have come to us from generations of scholars and investigators, which must be made a part of "what every child knows." This kind of knowledge acquisition requires considerable "mental work," presumably of the sort which keeps it organized and relates it to other knowledge. Children (of all ages, including college students) continually seek ways of avoiding this kind of mental work; that is to say, they tend to do their thinking in terms of already familiar knowledge rather than in terms which will "fix" new knowledge. A common rationalization of college students, for example, is that they prefer classes in which they "discuss issues" (using their already acquired incomplete knowledge) to courses of instruction which require them to "memorize information by rote" (which is a highly inaccurate way of characterizing the acquisition of new knowledge). This tendency to avoid "mental work," however, can be found in all of us, and is not confined to children or even to college students.

The basic means used to facilitate the acquisition of new knowledge by systems of instruction is to give the knowledge an externally imposed organization which will aid its retention. This may be considered one of the major uses to which scientific knowledge of learning and memory can be put. It is not possible here to review all of the suggestions which might be made in designing instruction for this purpose. I believe, however, that special attention should be paid to the ideas of hierarchical learning sequences by Gagné (1965a), or "advance organizers" by Ausubel (1960), and of the use of review questions by Rothkopf (1965).

Accordingly, the system for adult learning must also include a means of external organization of information to promote retention. Perhaps some adults will have achieved, by this age, some good efficient strategies for self-learning. For such fortunate people, the process of self-instruction can go on with the use of almost any medium—books, motion pictures, television programs. But many adults will not have achieved these strategies, because our school systems are not noted for encouraging their development. The system must then provide the external organization to materials for learning which will make their contents most readily remembered. It would seem to me that some of the principles of programmed instruction, those that have withstood the tests of time, could be of great usefulness in the design of such materials.

Productive thinking. The crowning achievement of a system for adult learning is to encourage and foster a variety of productive thinking activities on the part of the adults who participate in it. The first and most important requirement for such activities is that they be preceded by suitable mastery of relevant skills, and by the learning and retention of relevant knowledge. It is simply not efficient planning, for example, that a study of public speaking can be undertaken by individuals who cannot compose good English sentences; nor that a discussion of local law enforcement can be done by individuals who do not have knowledge of the structure and function of local government.

So long as this critical requirement is met, the system can incorporate a great variety of creative and productive-thinking activities. For example, there can be groups devoted to the development of artistic products of all sorts, from pottery to drama; groups whose function is to forumlate and test out action programs in civic and political affairs; and other groups which conduct seminars seeking to gain new insights and appreciations of literature, philosophy, and history. There would seem to be no special reasons for limiting the scope or nature of problems to be considered by well-informed adults. The key idea, again, is to avoid the operation of "courses" which, owing to the necessity of meeting the educational needs of everybody, regardless of his preparation, take the form of simplified chats from which virtually nothing can be learned.

Summary Statement

In summary, then, it would appear to me that the engineering logic of systems design and development could be successfully applied to the design of a system for adult learning. I have assumed that such a system should serve the needs of adults for continuing to learn, beyond the confines of formal degree programs. The basic logic, however, would apply to formal programs like college instruction if one wished to make the kinds of transformations necessary.

First, there must be a definitive determination of needs, which I conceive to be possible by a method which combines the information from adults themselves with projections of future trends made by knowledgeable "experts." From such a study, the purposes of the system as a whole could be defined. Once these goals are stated, system design can continue along lines that are relatively well known, to define subordinate objectives, allocate functions to both human and machine components, outline a plan for operation, and a set of techniques for field testing. In broad outline, I should expect the system to contain subsystems devoted to the functions of administration, instruction, student guidance, and evaluation. It is notable, perhaps, that the function of guidance in such a system is quite as essential as the functions of other subsystems. It would be a key factor in avoiding the aimlessness of current attempts at "adult education."

Each of these subsystems needs a careful design. I have made a few suggestions regarding the structure of an instructional subsystem. This subsystem must first of all attend to the need for "filling in" those basic skills, including numerical computation, tool-using, oral and written communication, that are known to be essential to the pursuit of learning itself. Secondly, it must insure the acquisition and retention of many kinds of previously codified knowledge. Selecting this knowledge is of course a difficult problem for system design to face. But the function should not be avoided or skipped, on the basis of any kind of rationalization (such as, for example, that "knowledge rapidly becomes obsolete").

Systematic knowledge, whether it is looked upon as "facts," "principles," or whatever, is essential to the successful performance of the third instructional function I have described, namely, creative activity and productive thinking. Such a function, in its many varieties, represents an ultimate goal for a system of adult learning.

REFERENCES

Ausubel, D. P. The use of advance organizers in the learning and retention of meaningful verbal material. Journal of Educational Psychology, 1960, 51, 267-272.

Briggs, L. J., Campeau, P. L., Gagné, R. M., and May, M. A. Instructional media: A procedure for the design of multi-media instruction. Pittsburgh, Pa.: American Institutes for Research, 1966.

Commission of the Professors of Adult Education. Adult education: A new imperative for our times. Washington, D.C.: Adult Education Association of the U.S.A., 1961.

Gagné, R. M. The conditions of learning. New York: Holt, Rinehart and Winston, 1965. (a).

Gagné, R. M. Educational objectives and human performance. In J. D. Krumboltz (Ed.), Learning and the educational process. Chicago: Rand McNally, 1965, pp. 1-24. (b).

Gardner, J. W. Self-renewal. New York: Harper and Row, 1963.

Hendrickson, A. Adult Education. In C. W. Harris (Ed.), Encyclopedia of educational research (3rd Ed.) New York: MacMillan, 1960, pp. 30-42.

Liveright, A. A. Lifelong education. Bulletin of the National Association of Secondary School Principals, 1964, 48, 99-110.

Luke, R. A., Bergevin, P., McKinley, J., Smith, R. M., Clark, R. C., Thiede, W., and Resley, R. F. Research within selected programs. Review of Educational Research, 1959, 29, 286-296.

National Education Association, Project on instruction. Schools for the sixties. New York: McGraw-Hill, 1964.

Rothkopf, E. Z. Some theoretical and experimental approaches to problems in written instruction. In J. D. Krumboltz (Ed.), Learning and the educational process. Chicago: Rand-McNally, 1965, pp. 193-221.

Sillars, R. Seeking common ground in adult education. Chicago: Adult Education Association, 1958.

Tyler, R. W. Some persistent questions on the defining of objectives. In C. M. Lindvall (Ed.), Defining educational objectives. Pittsburgh: University of Pittsburgh Press, 1964.

A SYSTEMS APPROACH TO ADULT LEARNING

Discussion by Earl I. Jones

Dr. Gagné's paper begins with a question which calls for a conclusion. The question is: "Can one apply the methods of system design and development to adult learning"? The conclusion given in the second sentence is: "Yes, indeed one can." And what can a discussant say about this diad? I agree with both the question and the answer.

A point worth making is that the same question and the same answer could apply to education in general, not just adult learning. In fact, it is difficult for me to imagine a man-made system, worthy of the appellation, where this would not be the case. It is probably true, though, that from a systems point of view adult education is more unsystematic than other "systems" such as elementary, secondary, or higher education.

Quickly summarizing the main points of Gagné's paper provides the following list.

1. A proper distinction is made between system design on the one hand and system analysis on the other.

2. Application of system analysis to the current situation in adult education would be less than sufficient. A broad approach of system design is required.

3. Application of system design requires: statement of purpose; identification of subordinate goals, functions and components; introduction of input and output concepts; and, in short, application of engineering logic.

4. Some denotative, connotative, and historical aspects of the problem in adult learning are given and then the paper proceeds to the subject of systems approach.

Up to this point in the paper there is really nothing with which I disagree, but there are at least two matters that bother me. The first involves the problem of identifying the goals and sub-goals of the system. Gagné, it seems to me, struggled valiently with this problem; he deprecates the use of experts, but sadly concludes that they are necessary; he recognizes the necessity of identifying the needs of adults which might be met by a properly designed learning system; describes the kinds of questions that should NOT be asked; and provides a basic assumption from which one can proceed to formulate questions which SHOULD be asked. The basic assumption is that learning means a change in capability.

The history of education may not be rife with attempts to establish goals. However, if we were to query the graduate student in education, I think we would find that he would remember several examples of attempts. So I might ask why are such nostalgic examples as the seven cardinal principles or the goals established by the Hoover Commission insufficient?

When we start talking about purposes, goals, sub-goals and objectives, we are really talking about levels in the abstractness or concreteness of the statements. General purposes or goals are of necessity abstract. Consequently, they at best serve only to indicate boundaries or limits. The problem of getting from goals to sub-goals to objectives and then to bite-size bits of learning is horrendous enough to be terrifying. But to quote Gagné out of context: "In designing a workable and effective system, questions like these must be faced squarely and answered honestly."

I feel that the principles presented by Gagné are applicable to designing a system of adult education, but that the process would necessarily have to be iterative. First, one time through would not be enough. In system design generally this is not the case——when you go through a procedure once you have a system— you do not have to repeat it. Secondly, roughly the same set of procedures would need to be applied at several subordinate levels in the system, each of which would require somewhat the same kind of logic.

Finally, although it is true that systems design could be applied to adult learning, one would have to have a sponsor who could provide both strong financial backing and the essential authority to implement the results. One of the biggest problems in putting instructional design or system design into effect is that it is so hard to break into the system to get what one desires implemented that sometimes it seems that the analysis has been a waste motion.

THE MEASUREMENT OF THE PRODUCTS OF TRAINING

Philip H. DuBois

It is obvious that the products of training consist of changes in behavior. As with many more or less psychological statements, this formulation is far too simple. It is the purpose of this paper to point out some of the complexities of the problem and tentatively to suggest a few steps toward developing answers. First of all, there are three kinds of behavior that must be considered. Training exists to develop changes in real-life situations. In Naval technical training, for example, the real kind of proficiency that we are interested in developing is skill in actual billets on shipboard or at shore station. Such skill might be developed, of course, in on-the-job training but in our system of class instruction we develop what might be called class proficiency, which, we expect will overlap the skills required in the real-life situation. Certainly one of the major tasks of the educator is to insure that as far as possible the skills developed under instruction are actually the ones that will be pertinent to the real situation.

There is a third kind of proficiency that enters the picture: test proficiency defined as the ability to perform in a limited situation in such a way that the scores are indicative of behavior in a larger situation. Hopefully test proficiency is indicative of skill in the real situation but conventionally we build our achievement tests around course content. Accordingly it is likely that they are generally more closely related to course proficiency than to skill in real life.

One of the reasons is, of course, the fact that real-life skills are often too complex and too interrelated with activities of others to be taught in a single course. Our course structure is primarily an attempt to fractionate the total skill into instructional units of such a magnitude that they can be handled in a given amount of time and with given facilities. It is the aim of a curriculum to develop the total abilities required in the real situation. The course is merely a convenient sub-division.

When we measure the results of training we should keep closely in mind what the measuring device is supposed to indicate. The parts of a test, generally items, are sometimes conceived to be a sample from a universe. We need to decide whether that universe is from the classroom, a micro-situation, or whether it is from the real world, the macro-situation.

Classroom measurement is merely a means to an end in the same way that the classroom is a means to an end. The tendency to focus achievement testing on instruction itself has a certain amount of convenience but it obscures its main objective, which is to increase real-life proficiency.

The role of the measurer in education is crucial. The power to test is the power to prescribe the curriculum and the power to determine the characteristics which are required to succeed. He who can write the examinations can determine the nature of instruction and ultimately modify the society into which the products of educational instruction enter. This role of an examination system has not always received the addition it deserves. For many generations the very fabric of Chinese society was determined in appreciable part by the competitive civil service examinations based upon the Confucian classics. Mandarin office-holders were literary scholars who were exceedingly influential in creating the relatively peaceful world the Chinese liked to enjoy.

It was not only the Mandarins who helped create this atmosphere. Would-be Mandarins, aspirants who never succeeded in passing the examinations, were far more numerous than the actual office-holders, since in any given examination the passing rate seemed to be around 3%. The great opportunities that were opened up by scholarship exhibited in competitive examinations made literary pursuits a way of life for an appreciable portion of the population.

Here in the Western world our examinations for certification and licensing have a somewhat similar role. It is hard to believe that our society could have achieved its present high standards in medicine, surgery, law, and other professions without high examining standards both in the schools and by subsequent official bodies. It is the pressure of the licensing and board examinations that forces schools to set up high standards of classroom achievement.

The military establishment has also found competitive examinations for promotion to be an important tool in the creation of technical competence. The extent to which tests can be made that actually reflect the status of the skill of the individual is the key to the success of a program of training and evaluation.

Of course there are great dangers in a testing program. Any testing system has built-in obsolescence. Ensconced examiners tend to pull old material from their files and, as time goes on, examining may become less and less pertinent. At the beginning of the twentieth century the Chinese found themselves with a governing class which had been selected largely on the basis of individual merit but the criteria used in developing the tests were those of an earlier day. The Mandarins knew little of science and little of the real nature of the threat from the West. Their decisions were in time so ineffective that the examination system itself was abolished in 1905.

I suspect that such effectiveness as the Chinese system had through the centuries was actually the result of two factors of which they were unaware: Spearman's _g_ (which he wrote about the year before the examinations were abolished) and the exceedingly low selection ratio in the Taylor-Russell sense. There have been times and jurisdictions in this country when our civil service examinations depended more on the measurement of _g_ and less on the measurement of specific skills required by the job. However, it should be pointed out that the United States Civil Service pioneered in the development of job sample tests. Back in the 1880's the Civil Service Commission deliberately studied characteristics of individuals successful in positions to be filled by examination and attempted to develop questions to measure those characteristics. One job sample test was for postal clerks and involved sorting a test deck of cards in accordance with the addresses written on them. This was approximately 80 years before the development of the zip code.

Other excellent work has, of course, been done in the measurement of proficiency. Walter Bingham was the chairman of a committee on the classification of personnel in the Army during World War I. The extensive series of trade tests developed and standardized at that time successfully differentiated among skill levels for occupations such as auto mechanics, electricians, sheet metal workers, and structural steelworkers. These trade tests constituted a key concept in the organization of the modern fighting force with its numerous specialties.

I think we are at the point now where achievement and proficiency testing can develop a somewhat more theoretical orientation than it has had up to this time. I think we should examine the inherent nature of our measures and see whether the correspondence between test structure and skill structure is good.

The difference between proficiency and achievement testing is largely one of scope. If the training necessary to develop real-life proficiency has been divided into units, each of the units can be measured by an achievement test. The measurement of the skill as a whole is by proficiency testing.

With that minor distinction in mind I shall henceforth use the terms more or less interchangeably. Ultimately I think we are always interested in the total skill; for convenience we are willing to measure parts of that skill. If our curriculum has been well made, each unit adds to the ultimate proficiency.

Another distinction between the two terms is sometimes made. Proficiency is thought to be related to skill whereas achievement is considered to be a matter of knowledge. Achievement tests have been constructed for the school subjects, history, social science, and the like whereas proficiency tests somehow carry the connotation of a complete skill, such as mastery of a foreign language. The distinction, however, is certainly not sufficiently clear-cut to warrant much attention.

It seems to me that the major distinction to be made in the area of proficiency testing is the matter of dimensions. Actually in achievement measurement the dimensions are: (1) extensiveness and (2) sequencing.

It can be demonstrated that if all the items of an achievement test measure a little bit of the same thing the result is a test with item correlations which fall into the Spearman type hierarchy. The rank of the matrix (with communalities in the diagonal) is one and there is a single underlying factor in this case measuring extensiveness.

There are undoubtedly a number of educational achievements which fall, reasonably well, into this pattern. In acquiring a foreign language it probably does not make much difference which nouns are acquired first. A test of noun vocabulary would be one index of language proficiency and it seems likely that the matrix of item intercorrelations would be a single factor.

The same situation would probably apply to history and other school subjects. Although we understand the present in terms of the past, it probably makes no difference whether we study English history before Greek history or American history before Roman. Chronological sequencing is obviously possible but only a minor amount of it seems to be necessary for understanding.

On the other hand, there are numerous instances where prerequisites are essential. Scalar algebra requires a knowledge of arithmetic and calculus requires scalar algebra. In flight training there are necessary prerequisite skills in the manipulation of controls before take-off and landing can be attempted.

Whenever skills or knowledge fall into a fixed sequence so that the higher level of performance includes something not appearing in the lower level, we would expect a quite different pattern of item intercorrelations. Instead of a single factor pattern in the Spearman sense, we anticipate a simplex in the Guttman sense. Now it is not the correlations that form a hierarchy but the items. As the number of steps in the sequencing increases, the rank of the matrix goes up. In the pure simplex in which a constant amount of new and different material is added with each item and all item scores are completely predictable from the total score, the rank seems to be n/2.

What does all this have to do with actual measurement in the training situation? I propose that it opens up a line of investigation hitherto neglected: a study of the curriculum through analysis of achievement tests, and the design of achievement tests to represent the structure of the skill to be taught.

I suspect that at the present time most of our achievement testing yields rank one item matrices. This could result from either of two conditions: that the total area is unsequenced or that the test itself deals only with a single level of achievement. Although, as I pointed out earlier, certain branches of mathematics are definitely sequenced, a test in differential calculus might easily yield a rank one matrix if the students taking the test were in a limited range of ability and if the items represented pretty much the same level of skill.

The time has arrived, it seems to me, to use the analysis of achievement testing as a method for studying training. I am reasonably confident that initial investigations would be disappointing. I suspect that we would find rank one matrices when we thought instructional material had been taught in a sequenced fashion and I think we might find high rank matrices when we had reason to believe that the training domain was flat. Such inconsistencies, however, could easily be a stimulus to re-examine the training itself. Eventually we should be able to develop new specifications for achievement testing closely tied to training objectives. As far as course-related tests are concerned they should yield low scores for the untrained, high scores for the trained, and intermediate scores for the partially trained. In addition we should have a way of relating internal test characteristics to the nature of the skill which is taught.

It seems to be likely that these characteristics would come out more clearly in a proficiency test, which attempts to measure a total skill than in a course-related achievement test. With proficiency testing the range of talent would tend to be higher and item correlations would probably be more variable. I should think that there would be a much better chance of finding, scattered in the item matrix, a Guttman scale or two.

Item difficulty, in the psychometric sense, is by no means synonomous with positioning in a hierarchical order of items. If there is a hierarchical order of items then, of course, the item difficulties will be similarly ordered. However, if a group of items have an identical level of difficulty they need not belong to the same hierarchy or any hierarchy at all, since the individuals passing could be very different and the item correlations need not be particularly high.

The approach I am suggesting is closely related to diagnostic testing, in which an attempt is made to find specific areas of proficiency and of disability. The focus in diagnostic testing is, however, on the individual rather than on the development of a training program.

As soon as we learn better how to build measuring devices and training procedures simultaneously, the analysis of item matrices should yield information on the two dimensions of training: extensiveness within a level; and ordering of levels.

For maximum usefulness identical measures would, of course, have to be administered to trainees in various stages of training. This need not interfere with the use of single level achievement testing within limited areas of training.

Various well-recognized methods exist to determine relationships between real-life skills and our measuring devices.

In the present context, it appears that we are concerned with something more than merely knowing whether our tests are related to proficiency. Obviously validity is important, but we could have a substantial relation between test and performance without the test reflecting the stages through which skill has developed. It seems to me that an attempt to relate test development more closely to curriculum development and to use the inter-relationships of items as cues to appropriate organization of educational content may have merit.

STRATEGIES OF ACHIEVEMENT TESTING

Winton H. Manning

Important objectives of the achievement test include: 1) assessing the degree to which an individual has acquired factual information in a field, 2) determining how well he understands principles and their application, and 3) measuring facility in solving problems germane to the subject matter. In the past, "psychometric" theory as applied to achievement testing has been dominated by theory arising from intelligence or aptitude testing. Recent developments in computer assisted instruction and programmed learning have, however, underscored differences between aptitude testing and achievement testing, which have existed all along, but which were largely neglected. A system for individualizing instruction allows each pupil to proceed at his own pace, with performance criteria that are identical for all pupils. Items for achievement testing in this case must be designed to indicate whether the desired behaviors have been mastered, rather than to discriminate among individuals. It is well known, however, that achievement tests are not constructed along this model; instead, items are selected to obtain maximum discrimination among individuals. Cox has described the dilemma by pointing to items which no one might pass at the beginning of instruction, yet all would pass at the end. This is the issue of criterion referenced versus norm referenced tests. It seems evident that some rethinking of psychometric practice may be expected to result from the computer sciences revolution in education.

Of course, the computer-based test raises other problems. The College Entrance Examination Board is supporting work to develop a computer-based test in physics, by which a student would be routed, depending upon his responses, to strings of items which are simplicially arranged to reflect the hierarchical structure of knowledge in the field.

How soon we may expect the 1,500,000 young people who take about 3,000,000 College Board tests each year to do so with direct communication links between computers at ETS and the campus, I cannot say, but it is a distinct possibility. The vistas of electronic communication are intriguing to contemplate, but I am convinced that for a long time to come, certainly within the life times of those in this room, we may expect students to continue to take tests which superficially at least are not so different in appearance from those they now take. However, let me underscore the word superficial.

This arises from the fact that our educational system, though it should be seamless, has and will continue to have major chasms, spanned by bridges which are significant realities. Specifically I refer to the process by which over 50 per cent of the products of 27,000 secondary schools distribute themselves to 2500 two- and four-year colleges, and other post-secondary educational institutions. This is the chasm across which the College Board has been building bridges for 67 years since Nicholas Murray Butler and Charles Eliot talked their conservative fellow educators into relinquishing a bit of their disordered autonomy and founded a central examining agency for college admissions.

What functions do achievement tests serve? I shall not exhaust the list by identifying four: 1) selection, 2) assessment of quality of educational programs, 3) a credential certifying a level of accomplishment, and 4) as a basis for placement in subsequent courses. All of these primarily serve the needs of institutions with secondary benefits accruing to the student. The logic of this is somewhat paternalistic.

Another function that achievement tests should serve is that of teaching. I do not refer to teaching subject matter, but to teaching the student about himself.

In order to teach students about their achievement in a field of study we should acquire a great deal of information about the nature of his understanding. Telling him that he did better than 43 per cent of some specified, often remote, group tells him something but not a great deal. To say that he mastered a given proportion of the total number of things he might have mastered, tells him something else, but still not a great deal. The obvious need is for a testing system which efficiently provides diagnosis; consider, for example, the placement questions confronting an entering student and his institution. One of my colleagues has catalogued these questions as follows:

Placement may refer to:

a) Selection (which student can take Russian)
b) Ordered articulation (algebra, trigonometry, analytical geometry)
c) Unordered articulation (which course in social studies)

d) Enrichment (ability grouping, acceleration)
e) Accreditation (knows chemistry, can take physics)
f) Vertical positioning in a homogeneous area.
g) Remedial placement
h) Combinations of (d) and (e), enrichment and accreditation
i) Grouping by mode of instruction
j) Free election

Conventional achievement tests cannot perform all these functions, nor is there statistical apparatus to provide the framework even if we had the tests. Furthermore, student and institutional collaboration seem indispensable unless we wish to foster an authoritarian shunting system, by which students are classified and routed mechanically. As an example, a test which reflects accomplishment in a secondary school subject may not contain the elements necessary for placement in a later course, unless there is excellent articulation across the chasm between school and college—and there is not.

A proper topic with which to start a discussion of strategies of achievement testing is the treatment of items. Tom Donlon, at ETS, is working on a College Board project dealing with distractor weighting. He has nicely summarized my first point as follows:

> "The two most popular techniques for scoring multiple choice tests are relatively simple functions of the responses: 'rights only' and 'rights with a correction for guessing.' The popularity is not unmerited. Both of these scores have repeatedly provided reliable and valid estimates of ability in many fields. They lend themselves to relatively simple procedures for processing data. Finally, their success has vindicated the logic which underlay their 'invention': unless the test construction is hopelessly faulty, the most able people are those who achieve a high 'rights' score and a low 'wrongs' score. Nevertheless, these scores may be criticized for a relatively inefficient use of the information which, at least in theory, a multiple-choice item may present.
>
> "This criticism does not arise from any radically different premises, but from a refinement of the basis on which the present techniques themselves are built. If it pays to single out the most correct response and give a positive credit for its choice, why not establish a hierarchy of correctness among all the responses and give differential credit depending on the degree of correctness? In other words, if we can consistently write key responses such that they are chosen by a more able group, why can't we write different kinds of wrong answers, such that the less correct the answer, the less able the group selecting it?"

I shall not summarize the evidence on this familiar proposal, except to say that Gulliksen pointed out that only negligible increments in reliability of a conventional test may be expected by distractor weighting. Fred Davis has done the most interesting empirical study—where for a 45-item mathematics test the parallel forms reliability was increased by an amount equal to a test fifty per cent longer. Short tests might therefore benefit from distractor weighting, especially if we knew better ways to write distractors of differing correctness.

Let me now turn to the problem of formula-scoring, namely the familiar (R - W/4) used widely in five-choice objective tests. Robert Boldt has recently completed another in a series of studies done at ETS suggesting that the random guessing model on which such corrections rest is incorrect. In his study, he utilized item responses of 103,000 SAT candidates to construct two tests—a difficult test and a broad range test. With such large samples a great number of scores fell at the chance level or below on the difficult test. The regression of these chance scores on the broad range test was linear and highly significant, although the same regression did not obtain throughout the entire range. This is consistent with a less well controlled study of the regression of chance and below scores on freshman grades, which suggested that with appropriate corrections for restrictions, validities were about the same for those above chance, i.e., about .50 and .60. These results cannot be explained on the basis of propensity to guess and it appears most appealing to me to invoke a concept of negative aptitude or achievement. There are people who are well educated in misinformation!

It seems fair to say that although no study of differential weighting has demonstrated that special weighting schemes are superior; nevertheless, strong interest in this possibility persists. This seems based on the view that variation in the amount and nature of incorrect choices could offer real possibilities for improving assessments of achievement and aptitude. Recently, Shuford, Albert, and Massengill stated:

> "Upon reflection it is quite apparent that all techniques in current use for assessing the present state of a student's knowledge fail to extract all of the potentially available information. In the case of objective testing...the response methods upon which they are based extract only a very small fraction of the information potentially available from each query."

On the other hand, Donlon has remarked that "wrong responses may be a kind of psychometric 'fool's gold' that relates to little except itself or, alternately, be such a low-grade ore that the labor of extracting it may exceed the rewards." Obviously, I do not share this latter belief, although I admit that there are grounds for argument.

Let me now refer you to a paper of Bruno de Finetti, which opens another avenue for increasing the information to be derived from a subject in a test.

How should an individual behave when confronting the problem of choosing among several alternatives. For someone who is certain as to the right answer the problem disappears; but usually there remains a greater or lesser degree of uncertainty. De Finetti takes the view that uncertainty is more common than we care to admit, and that the decision ought to rest upon the degree of belief—the subjective probability—that the individual entertains about the correctness of each of the alternatives. Contrary to this point of view most psychometric theory presupposes only two situations: complete information or complete ignorance, with complete ignorance ordinarily dealt with as if it were a blind guess. Although de Finetti's position is somewhat in the spirit of earlier work on confidence weighting, he unfolds in his brilliant paper the logic of evaluating responses to an item within the framework of subjective probability theory. Most certainly, for this system to work it is necessary that the subject be educated in these special methods of responding, and in risk-taking within the framework of decision theory. Validity in the sense de Finetti is talking about refers to the recovery of the subject's partial knowledge. The method is interesting because the task has changed.

Referring to the de Finetti scheme, Lord and Novich have stated: If we

> "Consider the expected value of the score s' with respect to the examinee's personal probability. It may be shown that this expected value is maximized when $r_j=p_j$ for $j=1,2,\ldots,A$. Suppose then that this scoring formula is used and suppose that all examinees understand its properties, know their p values and are able to make the necessary computations. Then if each examinee is motivated to respond so as to maximize his expected score, with respect to his own personal probability distribution this continuous scoring scheme will actually cause each examinee to reveal that distribution for each item."

Let me also refer to a rather obvious linkage which may be made between distractor weighting and de Finetti's method of examining partial knowledge.

In this application, the subject would be confronted with an item which is really in a game format. Alternatives having different weights would be assigned probabilities, and a payoff would be established based upon this matrix. Some years ago Luce and Raiffa observed that "decision-making theory may well provide the link for a unified conceptual structure" in the social sciences. I suspect this may be especially true in achievement testing, where such complex scoring becomes relatively simple in a computer era.

Let me observe finally that my interest in this strategy of achievement testing does not rest mainly upon the degree to which it will improve test reliability or validity per se, although these are still important. I do not, in fact, expect this to happen at all. Rather, its value is in two other realms:

1) Improving our understanding of the individual's comprehension and use of knowledge would permit us to teach him something about himself. If he learns enough about himself, he may be expected to make wiser decisions, and that may be really what education is about.

2) In his paper on learning with games which simulate the environment, Coleman observes that the greatest structural defect in our educational system is the mismatching of time. "The child is being taught for a future whose needs have not yet impressed themselves upon him." The value of games, which grow out of a decision theory approach, is that they "bring the future into the present, allowing the child to play roles in a large differentiated society of which he otherwise gets hardly a glimpse." In this sense then, building a test within a game model would make it more like the real world, rather than the artificial world of test makers.

If we could recast our strategy of achievement testing so as to take account of the partiality of knowledge, the condition of uncertainty in which decisions are made, and the degree to which society may value those decisions, we would have in our hands a powerful teaching instrument which is badly needed.

REFERENCES

Boldt, Robert F. Study of Linearity and Homoscedasticity of Test Scores in the Chance Range, College Board RDR -66-7-No.6, Princeton: Educational Testing Service, August, 1966.

Boocock, S.S. and Coleman, J.C. "Games with Simulated Environments in Learning," Sociology of Education, 1966, 39, 215-236.

Cameron, Ben F., Jr. "Academic Course Placement," Internal Document, College Entrance Examination Board, April 18, 1967.

Coombs, C.H., Milholland, J.E., and Womer, F.B. "The Assessment of Partial Knowledge," Educational and Psychological Measurement, 1956, 16, 13-37.

Cox, Richard C. "Achievement Testing in a Program of Individualized Instruction: Some Considerations," Newsletter of the Learning Research and Development Center, University of Pittsburgh, Vol. 4, No. 3, February 1967.

Davis, F.B. and Fifer, G. "The Effect on Test Reliability and Validity of Scoring Aptitude and Achievement Tests with Weights for Every Choice," Educational and Psychological Measurement, 1959, 19, 159-170.

Donlon, T. and Austin, N.W. Differential Weighting of Item Distractors in Data Sufficiency Items, College Board Research Project 141-05, Princeton: Educational Testing Service (in progress).

Dressel, P.L. and Schmidt, J. "Some Modifications of the Multiple Choice Item Format," Educational and Psychological Measurement, 1953, 13, 574-595.

de Finetti, Bruno "Foresight: Its Logical Laws, Its Subjective Sources" (Annales de l'Institut, Henry Poincare, 1937, Vol. 7) republished in H.E. Kyberg and H.E. Smokler (Eds.) Studies in Subjective Probability, New York: Wiley, 1964.

de Finetti, Bruno La Probabilita: guida nel pensare e nell'agire Serie Scientifica, Instituto Universitario di Scienze Sociali, Trento, 1965.

de Finetti, Bruno "Memorandum on Weighting Problems," Communication to W.H. Manning: College Entrance Examination Board, June 2,1966.

de Finetti, Bruno "Methods for Discriminating Levels of Partial Knowledge Concerning a Test Item," The British Journal of Mathematical and Statistical Psychology, 1965, 18, 87-123.

de Finetti, Bruno "Probabilita composte e teoria delle decisioni" Rendiconti de Matematica, 1964, 23, 128-134.

Gulliksen, H. Theory of Mental Tests, New York: Wiley, 1950.

Kogan, N. and Wallach, M.A. "Risk Taking as a Function of the Situation, the Person, and the Group" forthcoming in New Direction in Psychology III, New York: Holt,Rinehart & Winston, 1967.

Lord, F.M. and Novick, M.R. Chapter XIV, Mental Test Theory, Third Edition, Princeton: Educational Testing Service, 1966.

Luce, R.D. and Raiffe, H. Games and Decisions, New York: Wiley, 1957.

O'Day, E.F. A Feasibility Study of Test Format, College Board Research Project, San Diego State College, June 30, 1966.

Shuford, E., Albert, A., and Massengill, H.E. "Admissible Probability Measurement Procedures," Psychometrika, 1966, 31, 125-145.

Slakter, M.J. "Risk Taking on Objective Examinations," Journal of the American Educational Research Association, 1967, 4, 31-43.

STRATEGIES OF ACHIEVEMENT TESTING

Discussion by J. Roger Berkshire

When these conferences started about ten years ago, the emphasis was on methods of research in training. The fact that prior to that time no one had been able to show any very impressive differences between two training methods was attributed to the fact that we had no adequate way of measuring student gains. Consequently the great hope of the early conferences was that with the method of measuring residual gains we would have a tool that would show us when one teaching method was superior to another. At this point I said, "I doubt it." It seemed to me that much of the variance in learning is associated with individual differences and that what is left over is chiefly associated with motivation. What is left to be influenced by methods is very small. Most of the time there is no difference at all. If you do get a difference that is statistically significant, it will be because you have a large number of cases and it will not be practically significant. If you should get one that is practically significant, the difference in the groups will disappear at the next stage of training.

It is very often said that the reason we need to broaden and intensify education and training is because of the growing complexity of our civilization. I find this hard to accept. My grandfather's civilization was far more complex than mine. If he wanted to go some place he had to know how to take care of a horse, he had to build a buggy, he had to put steel rims around the wheels of the buggy, he had to know all sorts of craftsmanship that I don't have to know. If I want to go some place it is easy. To provide for my heat during the winter is no problem. It was a real trick for him. I submit that the reason we have dropouts is because it is apparent to them they do not need much information in order to survive. If I ask you to list the things that we absolutely need to teach kids for them to survive you would stop quickly. Our life is simple.

We went along for a while—but in the last few years there has developed evidence that training is important—far more important than we thought. It is important far earlier than most of us thought ten years ago. The extension of education to more and more people by going down in the age groups, by going up in the age groups, and by the very fact that the population is proliferating, is creating a different order of training problems than the one we talked about ten years ago.

It is not now a matter of the small differences in the quality of it; the question becomes one of "How in the world are we going to teach all of these people all of these things? There just is not that much money, there are not that many professors, certainly not that many master teachers." I think the biggest hope lies in something that Charles Thomas said. He pointed out that the highest rated master teacher is the one who arouses hostility in the students, who gets them involved by means of this hostility, and so motivates them. I submit then, that the greatest hope of our future is probably the teaching machine because of its infinite capacity to arouse hostility in students.

COMPUTER ASSISTED INSTRUCTION (CAI): SOME FACTS AND FANCIES

James J. Regan

It should be indicated at the outset that the point of view from which this paper was written is that of a manager. The author has had for the past several years a responsibility, largely administrative, for a developing program of CAI research and before that time a more direct involvement in both programmed instruction and early applications of the computer in the instructional process (although still from the management side of things). As a result, this discussion will take the form of an overview.

The computer has a role to play in a variety of personnel functions. It has been used as an automatic data processor for a number of these functions almost from its beginnings. A general view of the uses of computers in the personnel area may be found in the proceedings of an ONR symposium on automating personnel functions (1965). The present discussion will be concerned with the role of the computer in a particular personnel function—that of education. There are a number of uses to which the computer can be put in furthering educational goals. It can keep administrative track of the students, simulate the environment in which a student is trained, serve as an elaborate aid in calculation, and administer and score tests. However, its potentially most important function in education is tutorial, and it is this role that is considered here. For the purpose of this discussion the tutorial role will be referred to as Computer Assisted Instruction (CAI).

There is no educational research and development area changing as rapidly as is CAI. For this reason and because the information is available elsewhere [e.g. Hickey (1967) and Zinn (1966)] CAI status in the newsletter sense will not be exhaustively dealt with here. Instead, a brief selected description of the status of CAI will be presented under the headings of CAI centers, languages, programs, support, and economics. A second topic will be concerned with status in the more general sense of "where is CAI and where is it going." A final section, in deference to the subject of this conference, will be devoted to research issues and opportunities in CAI.

What Makes up the CAI World

Selected major centers of R&D [Hickey (1967) lists 20]

The Watson Research Center at Yorktown, New York, is responsible for some of the early work in CAI, developing programs in areas such as German and statistics. These programs have been implemented using an author language called COURSEWRITER. CAI systems using IBM equipment, programs, and languages developed at Yorktown have been installed experimentally at a number of universities. A new CAI system (the 1500) has been developed at another of IBM's locations, the one at Los Gatos, California. Dartmouth College has a CAI program using a GE computer and a student language called BASIC. By using this very simple, somewhat restricted language, a library of programs in 16 categories has been developed. The system at Dartmouth is accessed by a network of New England schools. The Systems Development Corporation at Santa Monica, California, using an author language called PLANIT, has developed among other programs one in statistics and one in counseling.

Bolt, Beranek and Newman of Cambridge, Massachusetts, is responsible for some very early work in CAI which was concerned with the identification of non-verbal sounds. They also developed the SOCRATIC system which has a number of special features allowing the student to engage in a limited form of conversation with the computer. A program in medical diagnosis has been developed as has one in which the subject solves a mystery. The SOCRATES program, initiated at the University of Illinois, has been used to develop instructional strategies which have been tested using academic course material. Also, at the University of Illinois, the Coordinated System Laboratory has developed a large scale CAI system called PLATO. An emphasis in this system has been on improving the quality and decreasing the cost of a student station. Plans now include a 2,000 station CAI center which can deal simultaneously with a variety of educational topics. The Learning Research and Development Center at the University of Pittsburgh includes a relatively recent CAI center. This center is concerned with the development of improved pedagogical languages and student computer interfaces. The computer based laboratory for learning and teaching at Stanford University is devoted to a large scale operational CAI program at one of the local primary school systems in which reading, arithmetic, and spelling are taught to grade school children.

Languages

To make CAI accessible to the educator and to the student it is necessary to develop languages for use with the computer which do not involve the complicated encoding and other limiting features of conventional computer languages. Languages which are available for the student to use in devising his own programs or otherwise to assist him include the TELCOMP language developed by Bolt, Beranek and Newman which functions in important ways as a computational aid. The BASIC language at Dartmouth, regarded as an algebraic language, does not provide for constructed verbal responses. Among the teacher author languages are COURSEWRITER which does provide for constructed responses; AUTHOR, the language used by the University of Illinois SOCRATES system; MENTOR, a Bolt, Beranek, and Newman so called conversational language; and PLANIT, a Systems Development Corporation CAI language. In PLANIT it is possible to select from five types of frames one in which to present the material and, within these constraints, to operate directly to computerize an instructional program.

Programs

Hickey has identified some 150 CAI programs as of March, 1967. A review of these programs by subject matter reveals that almost a third of them are in the mathematics area. Of the remainder, a quarter are in education and in educationally related topics and almost all of the balance in academic subjects such as physics, psychology, economics, statistics and chemistry. The length of these courses varies from three minutes to one semester courses meeting two hours a week. The amount of each of these courses which is on line with the computer also varies. Most of the CAI programs have been developed by Project Plato at the University of Illinois, Stanford University, and the several universities using IBM Yorktown facilities.

The impression one gets from reviewing the list of CAI programs is that most of them are fragmentary and that they cover a rather restricted area of subject matter. Restricted not from the point of view of the university setting in which most of them are generated, since they cover a broad range of college topics, but restricted from the point of view of applications outside of the university. For example, a variety of skills are taught by the services in addition to information acquisition. These, generally not represented by current programs, include troubleshooting, decision making, perceptual motor, and team skills.

Support

With many other technological developments in education, the Military Services have been in the forefront of CAI development. For example, the Office of Naval Research has been an important supporter of a number of the CAI centers and programs which have already been identified and a number of others which for reasons of time have not been included here. ONR has participated in the support of the PLATO project and the SOCRATES effort both at the University of Illinois, the CAI center developing at University of Pittsburgh, the work at Bolt, Beranek and Newman, and recently some of the work at Harvard University. In addition, it has been involved in a smaller way with some work being done on computer assisted explanation, on the use of computers in troubleshooting at the University of Southern California, and on educational strategies both at Stanford, and Systems Development Corporation. Finally, ONR has in the planning stage a major CAI development which goes by the acronym SLASH. SLASH is a large scale development center to include a library, an assist department, a department for software, one for hardware, and a section for studies. The Bureau of Naval Personnel is engaged in CAI research and development at its Training Research Laboratory in San Diego and has produced planning studies for CAI efforts both at San Diego and Mare Island, California. It is also developing a substantial computer based educational center at the U. S. Naval Academy. The Army is a somewhat more recent entry into the CAI development arena but even so has programs under way at HumRRO, Fort Benning and Fort Monmouth. The Air Force has been involved for some time in special applications of CAI at its Decision Sciences Laboratory, Bedford, Massachusetts, and more recently at the Training Research Division, Wright Field, Ohio.

The Bureau of Research of the Office of Education (HEW) has supported research efforts in CAI for several years with a number of its current projects going back to 1964. Of some 20 projects presently receiving OE support, 19 are with universities or state education departments. Thus, the focus of these projects is academic, being concerned with secondary and university education. Major projects receiving OE support include Penn State University (CAI in technical education), Stanford University (reading and arithmetic curriculum for use by culturally deprived children), and Florida State University (Science curriculum for grades 7, 8, and 9). The OE program represents in sum the largest investment of any agency supporting research and development in CAI.

Economics

A ratio of 1/40 of instructional to preparation time has been reported by Penn State. Other figures go up to 1/70 or 1/100. These figures are somewhat misleading since much of this preparation time is involved in organizing the material. This organizational time might also be expected in instructional programming (e.g. traditional programmed instruction) which does not involve the use of a computer. The time spent in coding or entering the material into the computer, although it varies, may represent but a small portion of the preparation time. PLANIT for example requires little of the author in terms of coding. COURSEWRITER calls for somewhat more. The choice of language (if one has a choice) is importantly related to the instructional strategy. BASIC for example does not permit any constructed responses. PLANIT has five types of frames, thus in some sense limiting the strategies to be employed. However, along with other features, it offers an effective compromise between flexibility in instructional strategy and ease of use. In any case, it is inaccurate to blame the computer entirely for the preparation time.

In addition, it can be expected that instructional programming will be more efficient with the use of a computer than it was when programmed instruction began. The computer can assist administratively and in on-line editing (Berkeley, of Information International, Cambridge, Mass., is working on this problem with ONR support). Finally, these development costs must be amortized over users and time. The preparation time when looked at in this light can be assigned in a variety of economically feasible ways.

A preliminary draft of a HumRRO report by Kopstein (1967) on the economics of CAI contains some interesting comparisons. The report cautions the reader throughout concerning assumptions in the data and in calculating indices. Given the assumptions (and without going into them here) Kopstein predicts that CAI will in ten years time cost roughly 1/2 as much per student hour as traditional instruction. The costs for higher education are roughly comparable now although for small school primary and secondary education CAI costs are presently ten times greater. All of the above comparisons assume that instructional effectiveness is equivalent—in many ways a conservative assumption.

So much for status. Where does CAI stand and where will or should it be going?

CAI: Some Observations and Suggestions

One can get a contradictory feeling about progress in CAI to date. On the one hand it appears that tremendous steps forward are being made. Activity is certainly increasing, at least when measured by publications. CAI is just over ten years old. As of January 1966 Hickey (1967) had assembled 100 documents relating to CAI. As of October 1966, 140 additional documents had been collected. Every month CAI is being inaugurated on a developmental basis in new locations. Yet many of the most recent installations resort to programs prepared early in the movement. The range of subject matter programmed for the computer has not increased materially. The student, for the most part, is still using the typewriter with all of its limitations to interface the computer. Nonetheless, there is a great deal going on and much that is new and improved is on the horizon (e.g. inexpensive and improved student terminals), but little of it is operationally evident.

What are the major problems to be solved before CAI becomes an operational reality? Mitzel (1966) lists five: the hardware-software gap; appropriate measures for evaluating CAI; time required to prepare a course of instruction; devising a "mix" between computer and instructor components in courses of instruction; and incompatible computer systems. Kopstein (1965) sees six research areas which require attention: (1) subject matter structure, (2) instructor-student informational coupling, (3) student-instructor informational coupling, (4) measurement of progress, (5) motivation, (6) differences. Most of those involved in CAI would provide similar lists. There is one important problem that is seldom explicitly stated. CAI requires a "critical mass" of significant proportions in terms of facilities, funds and talent (and not necessarily in that order). The fact is that massive resources are involved. It would be impossible on the basis of the fragmentary programming and instruction available to date (albeit considerable in the sum) to satisfy the Office of Education or anyone else, unequivocally and in cost-effectiveness terms that CAI is an efficient, economical educational procedure. Public and private agencies are cautious in fostering CAI development. On the other hand, massive resources have been martialled for an attempt to get to the moon and for fighting a war on poverty. Education is a key in meeting these national goals. There should therefore be a greater willingness than there is to finance a vigorous attack on what everyone agrees are the mounting problems faced by the educational community. Astronomical increases in cost are occuring just to keep up while accelerating amounts of technological information further complicate the problem. CAI is one potentially important way out of this difficulty.

There are several other important problems to be solved in CAI on which more explicit agreement has been reached. There are two pervasive characteristics of CAI as it exists today. The first, recognized by almost everyone, is the limitless potential for enriching the learning process represented by computer based education. The second, perhaps lost to those who are directly involved in its development, is the presently primitive nature of the student-computer relationship. An important goal for the future is the enrichment of this relationship. This enrichment can occur on several levels; variety of material available in the CAI mode and the options for presenting and responding to the instructional material can both be increased.

Another problem of concern to Hansen (1966), Mitzel (1966), and others, deals with the tendency in CAI to want to program on the computer every aspect of a given block of instruction. Serious effort should be given to discovering an efficient "mix" among the on-line use of the computer, traditional instruction, and the off-line use of the computer. In this latter employment the computer, in addition to keeping records of the student's progress, can select material for his review and study, and either print it our directly or send him to where it might be found.

A final problem which will only be identified here is that of author-student languages and instructional strategy. This is really a dual problem. If one is willing to restrict the computer to,for example, a programmed instruction type presentation, providing a language with which the author gets the material into the computer is a relatively straightforward matter. If, on the other hand, complex strategies and complex materials are to be used the language problem becomes monumental [cf Spolsky (1966)]. Various intense attacks on this problem are underway.

Research and CAI

Since the topic of this meeting is Research and Classroom Learning, it would be appropriate to make some comments concerning research and CAI. From the point of view of educational research the most important feature of CAI is probably the availability of adaptive instructional modes. Adaptivity as used here refers to the fact that the computer can present stimulus material to the student as a function not only of his most recent response but of some patterning of his earlier responses as well as on the basis of a host of aptitude, personality, and personal history characteristics. This individualizing of learning affords for example an opportunity to study the relationship of individual differences (in terms not only of aptitude but of such things as learning style) and methods of instruction.

Gentile (1967) has suggested that insufficient attention has been directed during the development of CAI to research issues occasioned by the use of the computer, and the more general issues associated with human learning. For example, he cites some misgivings Cronbach voiced over a decade ago about the advantages of individualizing instruction. That is to say about methods of tailoring instruction to individual differences. Cronbach suggested that we really know very little about what differences are related to what methods of instruction and he further suggested that, until we learn more, perhaps some plan carefully tailored to the group derived mean would be the most effective way to go about conducting a learning session.

The questions and cautions raised by people such as Gentile (1967) are real and they are important. However, there is an implication in these questions that a more research oriented CAI effort should be embarked upon and that the so far predominate development effort should be slowed down. Pursuing the development of CAI at this stage is admittedly something of an "act of faith." And it is an important act of faith because the costs, in various terms, incident on pursuing its development are so great. However, there is something of a dilemma posed by trying to increase research at the expense of development. No genuine assessment of the effects of introducing the computer in the educational process is possible without a significantly large amount of development. Piece-meal comparisons of learning methods using fragmentary programs of instruction are simply inadequate. Development must reach a point where longitudinal investigations are possible and where the computer can be used in ways that realize its potential. Studies where a complex strategy of learning a complex task can be evaluated in a setting that allows the student a range of responses begin to exercise this potential. Some observations concerning CAI research features and possibilities follow, assuming some R&D balance has been achieved.

Hartley (1966) reviewed about a hundred and twelve studies which purported to compare programmed and conventional instruction. Only six of these 112 met the following four minimum criteria: (1) program involved more than five instructor hours, (2) there were more than fifteen subjects per group, (3) the time to complete the program was reported, (4) pre and post results were reported. Of the six studies meeting these criteria all indicated that the program instruction was superior. Findings of this sort suggest that the research to be conducted on CAI should meet some fairly general, minimably acceptable ground rules. These rules can include those of the sort

Hartley used. Such additional ones as the following might also be useful: (a) some indication of the on-line/off-line ratio in a given course of instruction, (b) a description of the nature of the instructional strategies employed.

CAI seems to be avoiding the sort of comparative evaluations Hartley reviewed, with their attendant problems. There were innumerable studies, including some early ones with which the author was involved, in which a given block of material or course was taught via programmed instruction and compared with the same material presented conventionally. As you know most of these studies showed that the programmed material was learned more rapidly and that achievement at the end of the course was about the same. Occasional attempts at cost effectiveness analysis indicated that programmed instruction, to be economically feasible, had to be confined to those courses in which large numbers of students were receiving instruction.

The reasons that it seems desirable that CAI avoid this type of comparison are several. For one thing, there are fundamental questions of research design strategy in comparative analyses of the sort described above. For example, if an experimental evaluation shows "B" better than "A" the result may hinge on the fact that "A" stood lower in a population of A's than "B" stood in a population of B's. One might then have suggested, as I believe Skinner did, that the next task was to devise a better "A". Beyond this sort of consideration however, in CAI the question arises as to whether or not achievement as measured by end-of-class test scores is really an appropriate measure of effectiveness at all. Nitzel (1966) suggests that such achievement tests are in fact not appropriate since CAI in a fully realized tutorial mode is truly individualized instruction. Thus, it would seem more appropriate to measure effectiveness by how long it took the student to arrive at some point of mastery and/or how many repeats he required to reach this point. In any case, confrontations between CAI and traditional instruction are not at this point the most useful way to advance development. Licklider and others have advised against such confrontations on any terms suggesting that it is too early to evaluate the potential of CAI. An example often used here concerns the Wright Brothers' aircraft, which, if it had been compared with the Pony Express, might have lost the race. If this loss had been taken seriously there might have been no SST.

There are a number of other research opportunities afforded by CAI which time will permit us to mention only briefly. The U. S. Naval Academy with CAI support is concerned with using the computer not only in a tutorial mode, which we have characterized as the essential characteristic of CAI, but also in a broader educational management sense. The Academy will use the computer to control a variety of media (the multi-media approach) to be used in instructional process. Here the computer in addition to functioning tutorially selects various other educational devices for use during portions of a given course of instruction, assesses the student's progress for the instructor, and decides when the teacher himself should be called in.

Smallwood (1966) has suggested that the most important role (one that receives insufficient attention) that the computer can play in CAI is adaptive——the capacity of the computer to vary its presentation as a function of a variety of information that it has about the student. Smallwood further suggests that there are two kinds of adaptivity that should be exploited. The first is the one which has just been defined, the second is the capacity of the computer to change its rules over time as it learns about the effectiveness of previous adaptive procedures. There is fragmentary evidence available to indicate that a student will learn much more rapidly once he becomes aware of the fact that he is in control of the learning situation——a state of affairs rarely found in a typical learning situation. However, in the case of CAI, the opportunity does exist for the student to control the rate at which he is given information, the kind of information he is given, and when he shall be given it. It is quite possible once he recognizes these options——options for example which allow him to call for material already presented or to speed or pass by information he has already mastered——that he will acquire information a good deal more rapidly.

There is evidence in programmed instruction and of a more limited form in CAI to indicate that the relationship between intellectual aptitude and course achievement is lower in these methods of instruction than is conventionally the case. If these findings are substantiated it suggests a variety of possibilities including the possibility that CAI involves aptitudes not ordinarily measured and that individuals less trainable by ordinary classroom techniques might find CAI presentations more effective. Associated with this finding, or as an extension of it, is the possibility provided by CAI of assessing the consequences of various individual differences including personality and learning style differences for teaching strategies. This is an area of considerable practical as well as theoretical importance and the last word can hardly be written on it without a thorough investigation in a CAI setting. Finally, general observations might be made concerning the effect on learning research over a broad front that will result from the availability of computers.

As one looks at the possibilities identified above, it becomes clear that, in addition to the usual features of an automated research activity such as rapid data collection and analysis, the feasibility of exploring in some detail the effects of a variety of variables on individuals as well as groups is now open through CAI.

Conclusion

There are many reasons why educators and educational research people have misgivings about CAI both in its present stage and as it might develop. These reservations may be founded on a variety of operationally testable positions. Objections such as, "one cannot learn, retain, or transfer information," or that "student attitudes are negative" can be investigated and are useful in the course of evaluating CAI. However, the position frequently taken which has to do with the allegedly impersonal and mechanistic character of CAI is a value judgment which is not operationally useful. In fact, rather than view CAI as impersonal it can be said, as Christal (in Office of Naval Research 1965) has suggested, that one cannot afford not to give students the personal attention available to them through the use of computer based educational systems.

REFERENCES

Department of Defense - Office of Education - NSIA. Proceedings of the Engineering Systems for Education and Training Conference, 1966. pp 168.

Gentile, J. R., The First Generation of Computer-Assisted Instructional Systems: An Evaluative Review. AV Communication Review, 1967, 15, 23-53.

Hansen, D. N. Computer-Assisted Instruction. Unpublished Manuscript, 1966. pp 27.

Hartley, J. Research Report. New Education, 1966, 2, No. 1, (Abstract only; original not seen)

Hickey, A. E., and Newton, J. M., Computer-Assisted Instruction: A Survey of the Literature (2nd ed.) Newburyport, Mass.: Entelek Inc., 1967. pp viii + 75.

Kopstein, F. F., The Amplified Teacher: The Guidance of Human Learning through Controlling Functional Automata. Research Memorandum RM-65-15. Princeton, N.J.: Educational Testing Service, 1965. pp 21. (Abstract only; original not seen)

Kopstein, F. F., and Seidel, R. J., Computer Administered Instruction versus Traditionally Administered Instruction: Economics. 1967. pp v + 43. (Pre-publication draft)

Mitzel, H. E., and Brandon, G. L., Experimentation with Computer-Assisted Instruction in Technical Education. University Park, Pennsylvania: Pennsylvania State University, 1966. pp iv + 127.

Office of Naval Research. Proceedings of Inter-Service Conference on Automating Personnel Functions. ONR Symposium Report ACR-125, 1965. pp v + 154.

Smallwood, R. D. The Role of Computers in Educational Systems. (Text of a talk presented at the Stanford Interdisciplinary Computer Symposium, April 16, 1966) pp 6.

Spolsky, B. Some Problems of Computer-Based Instruction. Behavioral Science, 1966, 11, 487-496.

Zinn, K. L. List of Systems and Projects. (Interim draft of Section IV of a Review of Computer Assisted Instruction Systems) 1966. pp 21.

COMPUTER ASSISTED INSTRUCTION:

SOME FACTS AND FANCIES

Discussion by Robert Gagné

Regan's paper is a thoughtful and informative account of what is going on in this field. Having to comment on it has forced me to think about things that bother me about computer assisted instruction. I have always had nagging doubts about computer assisted instruction—not because I do not believe it is possible—obviously it is—but because it has not seemed to me to be terribly useful. Jim's remarks have not dispelled these doubts; in fact, I think that some of the things he mentioned serve to reinforce them. And it is not because I do not want to accept computer assisted instruction, but because I still do not understand its rationale.

Now let me see if I can catalog what I think to be the non-useful things about computer assisted instruction. The first is, as I said at a conference some years ago, that it really seems that what the computer is doing is page turning and that this is a very expensive way to turn pages.

We are told that the instructional programs to which the student is subjected are adaptive. This can mean several things. For one thing, it can mean that the instructional program is what some people call gaited in the sense that the individual can skip over parts of the program which he may already know. This is quite a reasonable thing. It seems to me, however, that in this case too I wonder whether it takes a computer to do this. The other notion of adaptiveness is the notion of learning styles. I still think that this is, at the present time, largely a myth. That is to say, it is something that people want to believe in but which they have very little evidence on which to base such a belief. The notions that some people react better to visual materials and some to auditory ones is an old idea that has a long history, with very little evidence to back it up. Cronbach's statement is that perhaps these kinds of learning styles do exist, but he could only find about one identifiable kind of learning style in the literature and this one pertained to a kind of personality reaction of students with a teacher. Therefore, it seems doubtful that could be represented in a computer.

The third point is that this kind of instruction is currently limited to a particular variety of instruction, a fixed reaction mold. The variety of material that is available on a random access basis may be large. Nevertheless, it does have fixed reaction with the student pressing a button or typing and the computer responding with a previously programmed response. It has been pointed out that this is really a very primitive mode of instruction and it may take many years to develop more sophisticated modes such as those which can be called tutorial mode, where the reaction to the student's response is really adaptive to what he needs at the moment, or other modes in which the student's thinking is, in fact, encouraged.

So, I think we are really very primitive in terms of what can be done.

Fourth, the technology of reacting to student responses is still pretty primitive. It seems to me that there would be a real breakthrough if the computer had a means of reacting to students' oral responding or even his written responding in a way which did not restrict him to pushing buttons. I do not know how far away we are from that; but technologically this would seem to be a complicated business which we are not approaching very fast.

A fifth point is that the whole matter of the coordination between the teacher and the computer, and the identification of their roles and of their relative functions is in a very primitive stage. We should think of the computer as freeing the teacher to do many things that may be more important than what the computer can do. I am not sure that it has showed signs of doing this. The teacher or somebody taking his place has to become involved in a way which simply supports the operation that is being carried out by the computer as opposed to being freed for other things. I do not need to remind you that there are many important teacher functions that apparently cannot be performed by the computer, such as motivating the student, and providing him with ideas, standards, and values. I simply remind you that there are educational functions that presumably no one has yet suggested that the computer can do.

Another point—and this a very important one—is that the emphasis tends to be on the machine and not on the student. I think it would be a bad mistake if we forgot that in any educational system the student is the central focus. Perhaps we need to reread Dewey on the child and the curriculum—I consider this the best thing that Dewey wrote on education. He wrote that you do not decide upon what to do about education by examining the curriculum, or by structuring the curriculum. You look at the child, you find out how he operates and that is the important question.

Now, taking all these things together I continue to have doubts that what will ultimately result from years of work, and the expending of enormous funds, will really be very useful. I think we must get back to the individual and I do not mean this in any mystical or even humanistic sense. Learning is fundamentally a highly idiosyncratic process. We can aid it by imposing external organization upon it, but we cannot make it occur that way. This is the nature of my doubt. Why must we take this tack at all? Why can't we build into the individual himself the capability of central processing which is fundamentally the event of learning. We have an enormous variety of information displays to depend upon. The individual can read books, he can look at pictures, he can listen to recorded messages. All of these can perhaps be improved for the purposes of stimulation or more broadly, communication. The computer can surely be of great aid in bringing about ease of communication in making available an enormous variety of sources of information, but none of these functions improves learning itself—they only improve the single condition of instruction that I have called the presentation of stimulant.

I believe that central processing can and should be built into the learner. The learner should learn to organize his own learning, he should learn strategies for selection, strategies for remembering, strategies for transfer. In summary, I think it likely that the computer is not going to replace the student's nervous system and therefore, it is bound to have limited usefulness no matter how hard it is pushed.

SIMULATION APPLIED TO EDUCATION

Edward R. Jones

Simulation as a training tool has evolved in 20 years to the point now where it is a major and essential part of the aerospace industry. It is needed to support space, aircraft, missile, and command and control systems so that appropriately trained personnel are available to operate and maintain them to the performance levels assumed in specifications and design. Two recent developments will markedly enhance its future utility:

a. Space Experience - The Mercury and most Gemini Astronauts were trained with simulation equipment alone to the point where they accomplished their first flights successfully (12). In addition, spacecraft simulators were interrelated with ground monitoring and control systems to form a systems trainer. Part-task devices, particularly for environmental and visual tasks, were used to supplement the major devices. The spacecraft program has truly been a demarcation point for simulation since it demonstrated dramatically the potential validity of simulators (5).

b. Commercial Airlines - The transition to jets and the accelerating growth in operations has required the airlines to use flight simulators for training even beyond the degree of reliance the military has placed on them. The SST, airport and airways crowding, and high aircraft operating costs have resulted in the economic need to accomplish as much airline training and proficiency measurements as possible in simulators (7).

This paper will examine how the technology now used extensively in the aerospace industry might be adapted to the classroom. The question that will be examined is the appropriateness of simulation for classroom training. My intent is to look more to the future than to the immediate situation because such speculation allows liberties not always available when one considers specific equipment capabilities.

This examination will be done from the perspective of an engineering psychologist who is both man and equipment oriented and who approaches problems such as this by synthesizing the factors related to the needs of the classroom, hardware capabilities, acceptance to the user, and cost. This results in a middle of the road position between that of the research psychologist and the engineer/computer specialist. We are particularly well aware of the problems of "gadgeteering", that is, equipment for equipment's sake without a clear analysis of its possible use based on need and utility, a predicament I wish to avoid.

Characteristics of a Simulator

Although a simulator can be broadly defined as any representation, model, or picture, its definition here will be restricted to a relatively accurate reproduction of a complex man-machine-environmental system with which man interacts in an operational or maintenance function (5). It is not open-looped as is a mockup, but is closed-looped and appropriately reactive to man's inputs.

It is either part or near-whole depending on the needs and economics of the situation. It contains enough of the essential elements to make it appear like the real situation. Generally, more than one person is trained at one time, but this depends on the characteristics of the system being duplicated. Training emphasis is on procedures, perceptual-motor performance, information processing, and decision-making. Environmental conditions such as motion and zero-gravity are occasionally added. The visual elements internal to the task (interior of the work station) tend to be duplicated more than exterior visual conditions such as airports, highways, and lunar surfaces.

A characteristic that distinguishes them from simple reproductions of the system itself is the capability to overlay malfunctions or out-of-tolerance conditions on the basic task. These conditions tend to be those rarely encountered or difficult to reproduce in the actual system (aircraft engine fire). Another key characteristic is the capability to measure man's performance for knowledge of results or administrative evaluation. In some cases, parts of a real system may be used in the simulation, but these are driven artificially or are related to an artificial environment. For example, the University of California at Los Angeles has a driving simulator which consists of an actual automobile on rollers with an artificial visual screen projected fore and aft. The military uses training problem generators to supply inputs to real systems. Occasionally, there are simulators for man himself such as that which reproduces his thermal output used for testing spacecraft under more realistic conditions. However, this paper will consider only simulators with man-machine interactions.

In summary, simulators are defined here as relatively exact reproduction of a part of near-whole task that generally duplicates man's interface with a complex machine or system. The device reacts dynamically with the individual by responding differentially to his inputs and allows out-of-tolerance conditions to be imposed upon the task. A simulator reproduces a specific, not a generalized task. For example, a duplication of the 707 crewstation would be a simulator, while a generalized duplication of the crewstation for any aircraft would be a trainer. It is generally assumed that transfer of training is achieved through this duplication. Although simulators are usually thought of as training devices, the aerospace industry uses them for other purposes as well. There can be a hierarchy of use sometimes for the same device which can include:

a. Research
b. Equipment Design
c. Man and Equipment Testing
d. Procedure Development
e. Training
f. Accident Investigation

For example, a zero-gravity crew transfer simulator used to demonstrate feasibility and for development of the techniques of going from one portion of a spacecraft to another will be used later as the training device for this task. The Gemini rendezvous and docking simulators used to develop engineering techniques and procedures were finally used for training. Later, I will refer to this concept of differential use for application to classroom learning.

Recent Developments That Enhance Utility

A number of recent developments relate to the potential use of simulators in the classroom. Some tend towards "gadgeteering," but others make the equipment more available and reliable, and add features that increase utility. I have always been leery of devices that are overly complex. Going back to my grade school and high school days, I remember that such simple electronic equipment as the classroom radio and the bells frequently malfunctioned. I wonder if anything much beyond this level of complexity might not be a maintenance nightmare for the classroom?

Some recent developments are:

a. Computer Availability - Digital computers are cheaper and are tailor-made to smaller applications. Remote access consoles and time-sharing capabilities add to their utility. Digital simulation has added flexibility because the design is not fixed initially and can be reprogrammed as new data become available. With analog devices, major design changes are needed to modify their characteristics. The hybrid simulator which combines analog and digital characteristics to the best advantage of both has also added to the flexibility of simulators. The use of a single central computer for satellite simulators or the shared use of a computer for administrative and training use has promise.

b. Flexibility - Building-block concepts have been introduced that permit the addition of features not anticipated or else not economically feasible at initial procurement. This allows one to start small and to build to complex systems as data become available and experience increases.

c. Performance Measurement - Scoring capabilities have improved because of the stability of digital devices and their computational capabilities. An example of an increase in measurement capability comes from one of the simulators used for research and development at McDonnell Douglas. Initially, it required six to eight weeks using film scoring techniques to obtain performance data on a five-minute run. Refinement of the device resulted in more reliable data on performance immediately after each run. Better measurement systems allow tailor-made training based on trainee proficiency.

d. Time Base and Difficulty Level - Special approaches to modifying the time scale of the training can increase its effectiveness. For example, SAMEX, used by the Army and Baylor University to train hospital administrators, duplicates for training in one to five days, two years of operation of a 450 bed hospital. Slow or fast-time techniques along with freezing a situation and replaying the preceding portion have utility in speeding the learning process. In addition, difficulty levels can be changed through an adaptive process making the problem increase in difficulty as performance increases. Adaptive systems are particularly useful for psychomotor tasks and information processing.

e. Software Availability - Probably most important for classroom learning is not only the simpler computer language and programs now available but the growing body of software that can be adapted for simulator use. These come from many sources including data and mathematical models generated in research and development activities. The availability of these data can reduce the cost of simulating complex situations. Related to this is the fact that many research and development simulators, as mentioned previously, are transitioned to training use.

Relation to Other Educational Techniques

An examination of the hierarchy of educational techniques suggests that simulation is one of the least abstract and close to real life situations (4). Table 1 is a gross classification of educational techniques and materials to give some indication of where simulation fits. The degree of feedback to the student as a function of his outputs generally increases as one goes down the list. For example, a lecture before a large class gives little interaction with the instructor, while a tutor is quite adaptive and functions very much as a closed-loop system

Table 1

CLASSIFICATION OF EDUCATIONAL TECHNIQUES AND MATERIALS

PERSONAL	AUDIOVISUAL AIDS	DUPLICATORS
Lectures	Written Materials	Mockups/Models/Specimens (in vitro)
Rote Responses	Pictures/Diagrams/Charts	Demonstrations
Questions	Audio Recordings & Playback	Games
Seminars/ Discussions	Still Projectors	Procedure Trainers
Tutors	Film Strips (Perceptual Training)	Simulators
	Motion Pictures and Television	Specimens (in vivo)
	ARC (Tab) Tests	Job Samples
	Teaching Machines	Intern/Apprenticeships
	Computer-Aided Instruction	

comparable to a simulator. The last column shows duplicators of real world situations beginning with mockups and models, which are nonreactive, and following through to intern-apprenticeships where the man is placed in the real situation and gets constant feedback based upon his performance.

The salient characteristic of simulators in the hierarchy is their very high cost, not only for development and production, but for operation and maintenance. This cost means that their application to the classroom situation will be limited to areas where their high potential payoff and considerable economic benefit can be derived. Even though an estimated $60-billion will be spent on education in 1967, the average yearly expenditure per pupil is, according to the National Education Association, only $564 in the elementary and secondary schools. Because of the diversity and number of governmental units and private groups that control these funds, there is no single major funding source for the considerable cost to develop complex simulators.

Present Applications in the Classroom

Simulators similar to those used for aerospace application are seen rarely in the classroom. The moot court is the best example of a high fidelity simulation in the classroom. But here people effectively duplicate the reactive elements, and thus there is no need for computers or controls and displays other than chairs, tables, and a bench. They are not necessary to simulate the essential elements of legal interchange.

One explanation for the limited use of simulators is that it is not the primary role of the elementary, secondary, or university systems to teach an individual to operate or maintain specific man-machine systems. This training is more appropriate for vocational schools and the military. Simulators for classroom use will take on a different form than those presently used in the aerospace industry.

Some Examples of Present Use

Examples of simulators presently used in the classroom are given below. They are all relatively simple, and their design and use have evolved gradually.

a. Driving Simulators - An inexpensive device used to familiarize students with the procedures and basic perceptual-motor skills required to operate an automobile (4). There are problems related to the utilization of driving simulators. One is that the real system is relatively cheap to produce even when equipped with dual controls. In contrast, a high fidelity driving simulator would be exceedingly expensive because of the need to reproduce a reactive external visual scene. A high fidelity driving simulator has not been built even for research in federal driving research laboratories because of technical and cost problems. The classroom driving simulators are open-loop devices that use motion pictures to duplicate the visual scene. They do have the capability for certain emergencies. There are 704 such devices in classrooms and they trained 250,000 students in 1966. In terms of transfer value, 12 hours in the simulator plus 3 hours in traffic is considered equivalent to 6 hours in traffic (11). That is, 12 hours in the simulator equals 3 in traffic.

b. Program Management Simulation Exercises for Defense Weapon System Management (USAF) - Exercises concerned with decision-making related to weapon system procurement have training attractiveness because the time-base can be speeded up tremendously in contrast to the life cycle of the real system (9). The future systems manager can select alternate designs and configurations based on cost and risk and then relate these to outside criteria of effectiveness. Thus, he can gain experience on a number of weapon systems in a short time period in contrast to the real situation where he would spend 5 to 10 years following a single system. These exercises are similar to management games used by the American Management Association, IBM, and several universities (2, 10).

c. Simulated Man for Medical Training (USC and Aerojet-General) - A manikin called "Sim One" is used for training residents in anesthesiology (3). Critical physiological parameters are displayed on a read-out panel. Cardiovascular sounds are present, and the manikin responds to drugs and anesthetics in critical areas such as pupil size. Hybrid computers are employed. Emergency situations such as heart arrest and fibrillation are possible. The mouth, throat, and other essential anatomical features are reproduced precisely. The device was developed under a $272,000 grant from the Office of Education. Some biomedical specialists feel that enough is known about some other body systems to simulate them in an equivalent manner.

d. Equivalent Response Learning (New York Institute of Technology)- Relatively simple devices train keyboard operators using a cuing technique that activates the keyboard to instill proper manual responses (9).

Considerations in the Transition from Aerospace to Classroom Applications

The first step in identifying where present aerospace technology might be adapted to the classroom should be based on a systems approach which is an organized effort to examine the classroom situation, identify relevant variables, and the desired output. These are then related to present and future technology, and the potential system is assessed finally on a cost/benefit basis. Such an approach is beyond the scope of this paper, but it probably would not identify many more areas for new application than those presently applied in the classroom because of:

a. Inappropriateness since much of what is taught in the classroom does not lend itself to simulation.

b. Very high cost which cannot be borne by the already overburdened educational system unless striking benefits are possible.

What are some of these benefits that must be realized? A simulator must be unique to justify its cost. Aircraft simulators were justified initially because they allowed realistic practice on emergencies too hazardous to reproduce in the air. Thus they had a unique capability that could save money in terms of accident reduction. Some factors that can be used to justify simulators are:

a. Consequence of Inadequate Performance - Improper performance is reflected either in economic loss or fatalities/ injuries, such as that by a commercial pilot.

b. Cost of Using Actual System for Training - The cost of operating the system such as an aircraft is high both in terms of revenue loss and salaries.

c. Hazard of Using Actual System for Training - The system in training use has an undesirable effect on society such as a nuclear power plant failure, a crashed commercial aircraft damaging civilian areas, or the annoyance of sonic booms.

d. Availability of Actual System for Training - The equipment is not available for training such as a spacecraft which must be flown correctly the first time or the transition of a pilot to a single-place aircraft where the first flight is without direct supervision.

e. Capability of Actual System to Provide Emergency Training - Out-of-tolerance conditions cannot be reproduced safely without damaging the real system or can be reproduced only at great expense, but can be reproduced readily in a simulator.

f. Requirement to Integrate Complex Team Activity - The integration of the activities of a large number of people in terms of information processing and decision-making is required particularly when complex information must be assembled and reacted on quickly.

g. Susceptibility of Task to Stress - Tasks which must be performed under stressful environmental conditions such as those involved in a space flight must be learned to a high level of proficiency to avoid disruption and this level is usually difficult to obtain on the real system because of the impracticality of repeating critical segments for training.

h. Capability to Centralize Training Location - The capability to reduce the number of training locations from many to a few with a simulator can reduce cost and achieve standardization.

i. Face Validity of Training - If trainee acceptance is necessary to assure their participation as with pilots or functional illiterates, then the "realism" of a simulator can be important.

j. National Need - The need for specialized skills to support a space program or medical technicians for developing nations can be important enough in terms of well-defined national goals that federal funding for simulation equipment might be required to speed up the training process.

Survey of Potential Areas

A variety of classification schemes might be used to examine the potential for simulators in the classroom. One might use perceptual-motor, procedural, decision-making, and personnel interaction skills for primary, secondary, university, graduate, professional, and vocational training.

With such a scheme, many areas for application can be dismissed quickly as inappropriate based upon a screening using the criteria of applicability. A summary of such a gross screening follows:

a. Basic Skills - There would seem to be little applicability to the initial acquisition of skills (6). Books, films, tapes, and programmed instruction are more than adequate as teaching aids.

b. Remedial for Basic Skills - Simulation-like devices might be used for perceptual training, but the applicability of true simulation devices would be limited unless face validity is necessary to motivate an adult population such as functional illiterates.

c. Handicapped Training - Handicapped persons, particularly those with performance problems, might be trained using simulation devices that duplicate some of the real tasks they must perform. An example might be driving a car where repetition training is needed on some task elements to overcome handicaps.

d. Vocational - The military has paved the way for the use of relatively low cost simulators, but there is a question of whether these are cost/effective for civilian training situations. Some areas that might be considered are TV repair, keyboard operators, and domestic and foreign medical technicians. The latter could meet the criteria of national need because of a political commitment. Other possible areas could involve the telephone information operator who, in the near future, will probably need computer assistance to search directories. Bell Telephone Laboratories has constructed an engineering simulation to investigate the feasibility of computer-assisted search which could be adapted for training use [redacted]

e. Professional - Law uses the moot court, and "Sim One" has been described as a specific technique for medical training. However, the area of greatest potential probably relates to training in medical diagnosis.

f. Operation of Specific Equipment - Railroad locomotives, nuclear power plants, power distribution systems, and chemical plants are all critical in terms of both safety and the economics of operation. Errors can have significant consequences. In the case of railroads, firemen are not available for the apprentice route formerly used to produce engineers. Simulators might be used to train a new engineer in the operation of the locomotive and in road familiarization.

g. Management Decision-Making - Applications might involve low to high-level economic decisions ranging from inventory control to the impact of a change in interest rates on the tax base. War gaming now uses computers to simulate the complexities of a military decision, either as a part of a complex command and control system as SAGE, or without reference to a specific command system. One advantage of management system simulations is that they can demonstrate cause-effect relationships to the student that might be masked in the real system because of its complexity. Further increases in the complexity of these simulations can be expected.

Future Possibilities for the Classroom

It is apparent that there are no obvious and exciting new areas for future application to the classroom. However, there is a common element that seems to permeate present and future use that bears examination. This relates to the development of integrative skills, and the reasoning goes something like this: individual skills and content such as mathematics, anatomy, and thermodynamics are taught on a part or isolated basis by conventional techniques. The application of these skills and knowledge is another matter. This is usually accomplished, sometimes haphazardly, and sometimes through apprenticeships or internships.

As an example, an engineer, despite the skills learned from university training, must spend many years developing the experience necessary to apply selectively to the design of the system information that he has. This process gained by experience allows him to make trade-offs within the technical constraints and relate these to schedule, cost, and performance. This skill in system integration seems to be based on experience rather than on what is taught in the classroom. Industry tends to measure an engineer's potential and salary, not on courses taken or his grades in the university, but rather on specific experience.

It is conceivable that another level of training can be achieved with classroom simulators which allow the student to develop the "experience" needed to solve actual problems. That is, to allow him to develop, while still in school, approaches derived from his testing alternatives in complex situations that not only integrate existing knowledge but allow him to experience the impact of schedule changes, failures, funding changes, and new requirements on the application of knowledge. Thus the simulation would allow experience which better approximates the conditions under which knowledge must be applied later. A similar kind of problem is seen with the new Ph.D. who, when asked to apply his research skills in an applied situation, is not flexible and experienced enough to identify critical issues and variables. His approach tends to be formula oriented and rigid rather than adaptive to the situation.

Gaining experience is more than reading or watching someone else do the job. It seems to require direct involvement in applying the information where feedback occurs regarding the consequence of the application. Simulators are ideal for this function since they allow the student to participate directly in near-real activities. For example, a number of aviation psychologists who are not pilots learned to fly complex aircraft simulators. Even though they had not flown a high performance aircraft, this experience allowed them to obtain a detailed understanding of the system and its characteristics way beyond that obtained from doing a task analysis, reading about flying, or watching someone else fly.

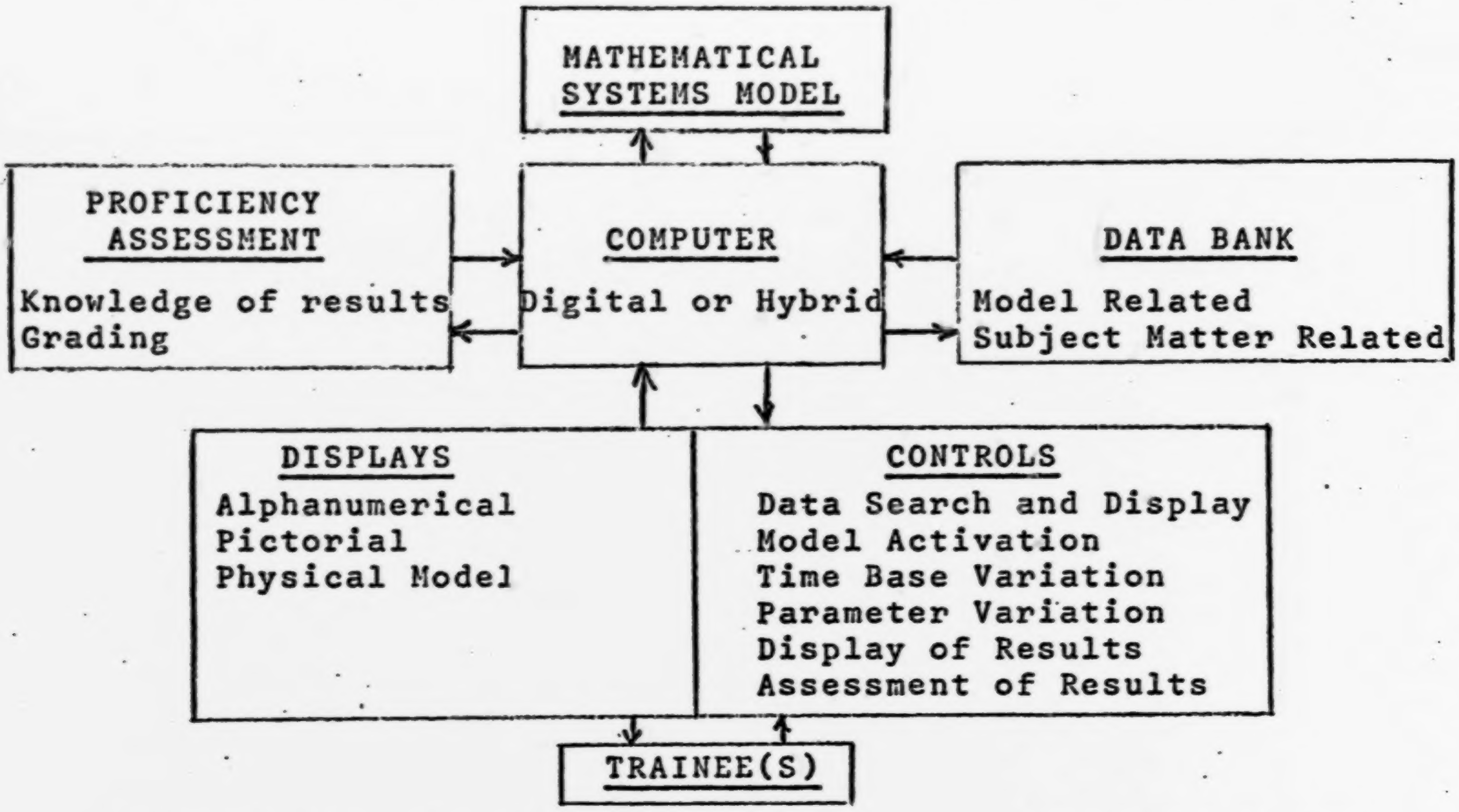
The Emerging Data and Technology

The major drawback to the applications just discussed relates to the cost particularly for software and system models. The potential for the classrooms becomes feasible if other funding sources develop the appropriate data and system models. Classroom simulation of complex functions becomes feasible as software and computer models are developed by others in a form that can be adapted without prohibitive expense. For example, in the space program automated laboratories and experiments are being developed, and physical, social, financial, biological, demographic, and personnel data banks are becoming more and more available. The point is that the data and computer models now used in many research programs can be adapted to classroom use allowing the student to gain fresh, firsthand knowledge of complex problems.

Some of the data and computer models in the past have been relatively simple and contained too few variables to be sensitive to critical interactions. But more complex models are now in use. For example, McDonnell Douglas has developed from empirical data a computer model based on SIMSCRIPT, a simulation programming language, of an emergency room in a specific hospital that allows assessment of the impact of variables such as mass emergencies, new facilities, and staffing changes (8). This kind of model could be used in the classroom to train hospital administrators to understand the dynamics of emergency room operation. An important aspect of any simulation for classroom use is the availability of measurement systems that allow feedback of performance for knowledge of results and grading purposes.

Some Examples of Future Use

If research and development models and supporting computer data are to be adapted to classroom use, then techniques are needed to allow the student access to the information. The following shows the gross characteristics of such a system:


MATHEMATICAL SYSTEMS MODEL
PROFICIENCY ASSESSMENT
Knowledge of results
Grading
COMPUTER
Digital or Hybrid
DATA BANK
Model Related
Subject Matter Related
DISPLAYS
Alphanumerical
Pictorial
Physical Model
CONTROLS
Data Search and Display
Model Activation
Time Base Variation
Parameter Variation
Display of Results
Assessment of Results
TRAINEE(S)

Essentially, the student would have a console that allows him to enter the system model, manipulate it, and obtain feedback of his actions either on a heuristic or preplanned basis. The student could test limits, introduce forcing functions, or determine sensitivities by introducing various conditions. Through this process, he could rapidly examine the interactions of complex information and determine the critical variables within a system. This process is analogous to the one used frequently in systems engineering. Several examples for application are:

a. Research - Training in research can be a wasteful process. An individual spends years, sometimes, before any feedback is obtained regarding the efficacy of his approach. Training should make him more sensitive early to research issues by allowing the simulation of a variety of problems. He could apply multiple approaches to data to test insights quickly and then assess all of these against "knowns" similar to those used by the chemist in teaching quantitative analysis. Such a process would train him in the strategies of research and allow him to quickly isolate critical variables. Practice using "automatic libraries" could complement such training.

b. Medicine - Sophisticated mathematical models are beginning to exist for systems of the body which give observable responses in physiological states when model parameters are varied. The fact that "Sim One" could be developed indicates such a capability. Students learning the systems of the body could gain experience through a simulation in the interactive effects that occur through disease when drugs are introduced. The pulmonary, cardiovascular, and renal systems are potentials for simulation. A more obvious area relates to the use of simulators to teach diagnosis.

c. Engineering - Engineering training tends to stress the acquisition of isolated skills and knowledge. Training in application occurs after the student leaves the university. Many times he is ill-equipped for considering the complex interactions that may occur within the elements of his technology as well as their impact on the social and economic structure of society. Computer models could allow these interactions to be defined better. For example, one could assess the impact of a new highway not only on traffic flow but on the area's economic and social structure.

CONCLUSIONS

Present and projected capabilities in aerospace simulation were examined to determine what aspects might be adapted for classroom use. The attributes of simulators as well as the criteria that must be met before they can be adopted were considered.

Simulators are expensive to acquire and operate. Their justification must be unique. It appears that beyond a few conventional uses such as for driver training that they could be applied best to the integration of complex skills and knowledge. This process would be relatively economical if existing computer models from research and development programs are adapted to classroom use.

REFERENCES

1. Allen, J., Operator and User in the Commercial Telephone System, Paper presented at Symposium on Human Performance in Systems Effectiveness, Naval Material Command and National Academy of Engineering, Washington, D. C., 17-18, January 1967.

2. Carter, L. F., Exercising the Executive Decision-Making Function in Large Systems in Training Research in Education, R. Glaser, Ed., University of Pittsburgh Press, 1962, 409-427.

3. Computer-Controlled Manikin Developed by Aerojet, USC, Technology Week, Vol. 20, 3 April, 1967.

4. Erickson, C. W. H., Fundamentals of Teaching with Audio-Visual Technology, Macmillan, New York, 1965.

5. Fraser, T. M., Philosophy of Simulation in Man-machine Space Mission System, NASA SP-102, Washington, D. C., 1966.

6. Gagne, R. M., Simulators in Training Research in Education, R. Glaser, Ed., University of Pittsburgh, 1962, 223-246.

7. Jones, E. R., Future Simulation Needs, International Air Transport Association, Paris, November 1966.

8. Kirkpatrick, D. C. and L. H. Whitehair, Computer Simulation of an Emergency Room Facility, McDonnell Report F458, 1 April 1967.

9. Proceedings of the Engineering Systems for Education and Training Conference, NSIA, Washington, D. C., 14-15, June 1966.

10. Smode, A. F., Recent Developments in Training Problems, and Training Research Methodology in Training Research in Education, R. Glaser, Ed., University of Pittsburgh Press, 1962, 429-495.

11. Stack, H. J. Driver Education Growing, but Most Youths Get No Courses, The New York Times, 2 April 1967,

12. Swenson, L. S., J. M. Grimwood, and C. C. Alexander, This New Ocean - A History of Project Mercury, NASA SP-4201, Washington, D. C., 1966.

SIMULATION APPLIED TO EDUCATION

Discussion by James M. Vanderplas

While I stand in awe at the munificence and magnificence of simulator systems, I share Dr. Jones' doubts as to their appropriateness for widespread application to the classroom. Dr. Jones has provided an excellent overview of the military and industrial uses of simulators. He has given us a fine set of criteria to justify simulators, as well as a survey of areas of their potential use. He has also shown that simulators are most applicable to teaching in areas requiring complex integration of skills or application of existing knowledge in such areas as research, medicine or engineering.

I would, however, like to modify his definition of a simulator. Recall that his definition includes the notion of fidelity of reproduction of the stimuli involved in the situation. I was a member of a conference involving Randall M. Hanes, Neal R. Bartlett, John W. Gebhard, and Anthony Debons in 1962 when we were concerned with a project involving simulational facilities for the study of displays for command and control systems. After wrestling for several days with the philosophy underlying simulational facilities, we finally decided that the entire project might well hinge on a definition of a simulator. Because the term simulator, we said, can be used to cover anything from a field test to a laboratory apparatus for maintaining an isolated factor in a problem, the following definition was adopted: "A simulator is a device or process for producing as simply as possible all the elements of a situation which are relevant for a given purpose." This covers the ground quite broadly but also imposes some rather severe restrictions.

It was our view that the utilization of simulators held promise for research on many of the problems of developing displays for the control of forces. In the concept of simulation, however, there is no attempt to include the entire situation. Only those elements which are deemed especially relevant after careful study of the situation would be included. Realism may not be necessary; indeed, in some cases it may even act to the detriment of the training for which the simulator is designed. In a paper I gave at a conference having to do with problems in the teaching of mathematics I was able to find a number of instances in which realism in teaching devices tended to retard rather than to accelerate the learning of mathematical concepts.

Another point has to do with the construction and maintenance of simulators. Dr. Jones says that simulators cost a lot, must be unique and are difficult to maintain. I agree, but I would point out that these are generalizations which arise in an aerospace context and which need not necessarily apply to classroom simulators. As I have defined it, a simulator would probably contain only those elements which are especially relevant to the subject matter under consideration. In simulating a grocery store activity, for example, as Dr. George Kreezer has done with his school for the educable retarded in the Psychology Department at Washington University, the cans and bottles do not need to be full of real food. One would not need to have real meat in the simulation of a butcher store. In simulating first aid activity, for example, one would not need actually to break the leg of the "victim." Simulators were originally designed to duplicate equipment; to include variables in a complex environment not available by other means; to feed back complex information; and to provide training for emergencies which, in vivo, would result in catastrophies.

Later, simulators were used as training devices. This evolution, it seems to me, tended to emphasize face validity as contrasted with other kinds of validity: concurrent, predictive, and construct. In World War II Air Force paper-and-pencil tests were found to be as effective as psychomotor tests such as the complex coordinator and the rudder control test in predicting pilot proficiency.

As far as I know there is no systematic study to determine the validity of simulators, as compared with other training devices, in classroom situations. I would emphasize that this needs to be done. I would therefore caution against the direct application of the aerospace concept of simulation to the classroom. Just as the principles of conditioning were applied without modification to classroom teaching, an unmodified application of simulational concepts to classroom work might result in a similar disaster. I remember that I was taught writing by the so-called Palmer Method, with endless repetition of making ovals, making sure that my fingernail rested on the paper and that the ball of my arm was used to move the pencil. But I never recall being reinforced for a correct response, and the teacher completely neglected the fact that each movement was an extinction trial.

Finally, I will suggest some augmentations of the use of simulators. Using the concept of a device simulator, it should be possible, with caution, to design classroom simulators which include the structural elements necessary for teaching certain subjects. Examples of such structural simulators would include automobile mockups already mentioned, computer hardware, toys, tools, action games such as Combat, Detective, or Medic, in which simulated weapons, surgical devices and prosthetics are employed to include realistic objects necessary for the task. Hospital staff disaster training, in which the actual equipment such as stretchers and ambulances, are manipulated in real time would be another example of this kind of simulator. By using the concept of a process simulator it should also be possible to design simulators which would include functional elements in the absence of realistic appearing structures. Examples here would include the moot court, not requiring the actual courtroom setup, but requiring only benches, tables, and perhaps chairs; business games; and diagnosis and other hospital procedures involving concepts and manipulation of data.

I also envision the design of hybrid simulators in which both structural and functional elements are included, but again with caution and with the need for relevance emphasized. Role playing in therapeutic situations is an example of such a hybrid system; managerial grid training; sensitivity training; and certain aspects of first aid training are also examples. Simulation of stock market operations using blackboard and chalk, rather than computer transmission devices; political system simulation; and war games would also be examples.

THE NEW LOOK IN EDUCATIONAL TECHNOLOGY IN THE PUBLIC SCHOOLS

James M. Dunlap

In considering the new look in educational technology, my first reflection focuses on the vast capability of computer processing in research. Four years ago the University City Schools received a Ford grant of $267,000 for innovation and research for a three-year period. This year we have several grants from the U. S. Office of Education. Although the University City schools have long been interested in research, data processing was slow and manually operated. The funds from outside agencies have speeded the process with the result that, each successive year, I believe we do a better job of evaluating how well we are teaching and how well the youngsters are learning. Research, of course, is particularly valuable in predicting outcomes. We are just beginning to examine how scores in prekindergarten, kindergarten, or third grade predict what a youngster may be doing later on in school or college.

A second very important use of computers is test scoring and reporting. Electronic processing certainly is time saving; teachers like it, and have learned to interpret computer results with understanding.

Another area in which we have been using computers in recent years is in student scheduling. In the high school we are making a beginning. The only way one can schedule 2000 students efficiently in individual programs is by computer. Other obvious uses of computers in schools are in the business and personnel offices.

As teaching tools, I have had mixed feelings about the use of the computer. I can accept the computer wholeheartedly as a tutorial device where facts or basic skills are taught by rote. I have less enthusiasm for the use of computers where ideas, concepts, judgments, and differences of opinion are being fostered, perhaps because of the difficulty in preparing adequate material.

The first time I heard that a computer could be programmed to counsel students I was horrified. And yet, when I think of a beginning counselor even after four years of undergraduate work and two years in a masters program, I sometimes wonder whether he is quite experienced enough to counsel a high school senior no matter how good the training. Now I am able to imagine that ten or twenty of the country's best counselors working for a year could prepare sound solutions to various problems a student might raise. Such computer counseling conceivably could be as effective, or [redacted] more effective, than counseling by a self-trained new counselor of limited experience.

THE NEW LOOK IN EDUCATIONAL TECHNOLOGY IN TECHNICAL TRAINING

G. Douglas Mayo

The trend, or new look in educational technology in technical training, seems to me to be unusually clear, considering the early stage of its development. This trend is away from mass, lock-step training and toward individually adapted training, carefully geared to performance requirements. When we move from this general statement to specific details, the projection into the future becomes less clear. But even so, there is a principle which tends to determine the direction of movement. Given an acceptable quality of training, the economics of the situation constantly maneuvers technical training toward the most efficient means of achieving the aims of the training involved. This stems to a large degree from the fact that the recipient of technical training, whether in the military or in industry, is paid to take the training. He is there because management wants him to perform tasks that he could not perform prior to training. Understandably, management wants him to learn the desired skills and to begin performing them as soon as possible.

Other factors will influence the direction of technical training to a greater or lesser extent from time to time, but the primary factor over the longer term is primarily economic. In using this term I do not imply a "penny wise" point of view. I use the term to connote simply, "the effective allocation of limited resources." Thus, the commitment of substantial funds to develop and implement an educational technology ordinarily is not a problem if it appears that the goals of training can be achieved more efficiently.

In attempting to project the future of educational technology in technical training my views will be guided by the economic principle stated above, or if you prefer, assumption. In addition I shall limit my projection into the future to a typical military technical training situation, although I suspect that most technical training will be affected by essentially the same forces.

Individually adapted training, which doubtless will be machine facilitated, will be adaptive in two respects: first, in terms of trainee aptitude, interests, and other pertinent characteristics; and secondly, in terms of the specific terminal performance that is desired——by management. In the first instance the instructional system is likely to be similar to computer assisted instruction in non-technical training situations. I expect it to reach a mature state through a series of steps which will include initial use of the computer largely to accomplish the routing of the individual trainee through training, beginning with his initial level of knowledge and skill and proceeding to the desired level in as nearly an optimal manner as can be accomplished.

This optimal routing will require a great deal of development effort in each instructional system. Initially much of the instructional materials, such as programmed booklets, audio recordings, films, and video tapes, are likely to be structurally independent of the computer. This is due to technical and economic reasons. Gradually the instructional materials will be structurally integrated into the computer based system.

In the second instance, which involves tailoring training to the performance requirements of the specific job which the trainee will fill, a number of changes from our present procedures are foreseen. For example, during the recruit training period the specific position in a given squadron to which the man will be assigned will be determined. The specific performance requirements of the position will be determined and the programming of the man's training to meet these requirements will begin. The trainee will proceed as rapidly as his aptitude and motivation will permit through the required training, to the end of his individual course, and to the position for which he was selected.

Suppose the trainee is not intrinsically motivated. Suppose he is content to make a "career of training" in a course that normally should be completed in a few weeks. There will be time standards at three levels: minimum, standard, and superior. Present corrective procedures, such as conferences, evening sessions, and termination of training, will continue in the case of failure to meet the minimum level of performance. A positive incentive to achieve beyond the standard time requirement will be included in the form of a monetary reward for rapid achievement of the objectives of the course. This reward will be based upon a percentage of the amount saved by early completion of the course. Trainees will be permitted to work longer hours than those of the standard work day if they wish to do so and will be rewarded for early completion of the course.

The units of which the course for each trainee will be constructed will be determined by "specific behavioral objective" or "terminal performance specifications," as is currently true in programmed instruction. These objectives will be assembled for each individual in response to the question of what tasks must be performed in the position to which the man will be assigned.

There will be no instructors as we now know them. There will be three primary instructional tasks. The first will be preparation (including validation) of adaptive material to meet specific behavioral objectives. These materials will include extensive branching and remedial loops. The second instructional task will involve selecting the behavioral objectives which pertain to the position for which the man was selected. It is at this point that the content of the course is tailored to the requirements of the position. Much of the information required to make a decision concerning what the content of each course should be, will be stored on discs and available from the computer. The third instructional task will be the management and monitoring of the training situation. The monitoring function will be accomplished in much the same way that automated production machines are monitored at the present time, although a "human touch" will be included in the management function.

The extent to which it will be economically feasible to decentralize training is not entirely clear. Progress made in economical communication from the computer to remote locations will be a factor. However, a larger factor in technical training will be the duplication of costly training equipment. Until relatively inexpensive, but effective, substitutes for such equipment are developed, technical training doubtless will continue to be centralized in training centers.

I have attempted briefly to point to some of the changes that may accompany the development of new educational technology in technical training. Obviously most of the detailed projections are largely speculation. One point that I feel is clear, however, is that the rate of change which has been so slow in the past as to be almost imperceptible, will be greatly accelerated in the future.

THE NEW LOOK IN EDUCATIONAL TECHNOLOGY IN COLLEGES AND UNIVERSITIES

Marion E. Bunch

The most important goals in educational technology at the college and university level are to improve the quality of the education and to structure the situations so that a larger number of qualified students can be accommodated without reduction in the quality of the education.

For many years the experimenter studying learning in the laboratory influenced the process of learning only in so far as he set the problem to be learned and the conditions of the study. As soon as the subject began the learning trials, the experimenter put himself in the background and became an observer, either recording the responses of the learner, the number and type of errors made, etc., or monitoring the equipment which recorded the behavior automatically. In the early 1930's, however, psychologists designed a number of research studies to investigate the effect of intentional attempts on the part of the experimenter to facilitate and direct the course of learning while it was going on. Several laboratory studies were carried out which were concerned with determining the effect upon learning of different kinds and amounts of guidance. These studies of the influence of tuition upon learning while constituting basic research on the process of teaching, were primarily the work of a few experimenters and did little more than call attention to the fact that the very real and complicated process of teaching is capable of scientific investigation.

In any list of the forms of aid that help the learner solve a problem, or gain greater understanding, or improve his performance, or acquire learning sets, we must include the teacher as providing the central tuition through questions, suggestions, instructional and informational guidance, as well as motivation and immediate rewards. The student can make extensive use of library, laboratory, and field facilities, all of which substantially promote learning.

Many studies have indicated that in the acquisition of skills, the more the trials of practice require the same kind of performance that will be required later, the greater is the likelihood the practice trials will be beneficial.

The principle may apply in graduate education where we recognize clearly that learning is an individual process and that individualized instruction is needed. Perhaps in the future when better and more exhaustive analyses of complex learning processes are available it will be possible to set up programs or computer assisted instruction in many specialized areas at the graduate level. However, there are exceptions where this does not seem to be possible at present. For example, a number of graduate departments of anthropology decided several years ago that they would not grant a Ph.D. degree in anthropology without field training the fourth year. There is nothing in the way of local instruction that can take the place of this kind of individual learning experience. This requirement became a rather severe one because most universities cannot support a student in the field for the desired research training in the fourth year. However, there are now Federal grants of $10,000 to $12,000 to support students in anthropology in field research in the fourth year of the Ph.D. program. In the physical sciences a piece of equipment might run many times such a figure without the cost being considered out of line. However, an expensive apparatus may be used not only by the student for whom it was purchased, but also by those who come along thereafter. Field research is necessarily individualized and it is hard to see how it can be simulated.

Let us take another instance of individualized learning. After an adult has finished his formal education he has learned more than solutions to particular problems and has acquired more than certain specific skills. Presumably, he has better habits in attacking problems. There are several studies of learning that are pertinent. For example, one finding is that in solving complex problems the degree of variability that the learner brings to bear on the task is very important. In the long run, the greater variability the learner shows in attacking a problem, the greater is the likelihood he will find a solution. On the other hand, if he approaches his task with a small behavior repertoire, he will exhaust his possibilities very quickly. So, one of the ways to instruct students to improve their ability in solving problems is: "If you don't see the solution quickly, the thing to do is to focus attention upon various alternative possibilities and find as many tentative solutions as you possibly can." In group learning situations involving selected complex problems it has also been found that a small group will often come up with a better solution than the best individual could find by himself. As you know, it was argued for a long time that a group solving a problem would not do better than the best individual in the group if just let alone; the presence of a group only meant that the best individual had to explain things to the others and point out why their solutions were not correct. However, it has been discovered that the group, in some learning situations, tremendously increases the range of alternative suggestions that need to be considered to solve the problem.

Now, the question is how to encourage this aspect in problem solving. Frequently in programmed instruction we want the responses highly structured and we do not want the learner to make too many errors. In fact, in many cases we may have written the program in such a way that when the learner is challenged to make a response he is likely to write in the correct one. It is argued this is good because the learner gets practice in doing the correct thing, but other studies show that this is not always the superior method.

We have several forms of aids that would vary in utility with the nature of the problem: laboratories, field studies, libraries, as well as programmed instruction and computer assisted instruction.

One thing that the teacher can do is to analyze the subject matter of the course with special attention to the level of conceptualization that is required for mastery. He then can try to make the steps small enough that they can be programmed. There are many instances in which well-accepted notions about the nature and size of the preceding steps have turned out to be subject to modification. While geometry is customarily taught in the tenth grade, some schools report they are teaching it in third grade and that the students are doing quite well. A few years ago the mathematicians on one campus stated that it was undesirable to teach calculus at the sophomore year. The reason stated was that in first year mathematics, time was taken up primarily in teaching students the mathematics that they should have learned in high school. Accordingly, the second year was needed to teach them the basic work in algebra before they could take calculus. However, the possibility of shifting calculus to the junior year caused much concern because other courses had calculus as a prerequisite—physical chemistry, for example. If calculus were shifted to the junior year, students would not be able to take physical chemistry until the senior year.

Physical chemistry is a prerequisite to other courses which students would not be able to take until after they graduated! The argument became quite heated. One thing that helped to resolve it was an article which appeared in a current issue of Science, calling attention to the fact that in England calculus was taught in the seventh grade. The level of conceptualization of the course in England was not known but it furnished some ammunition to the people on the campus who thought there was no necessity to shift calculus from the sophomore year to the junior year. Perhaps too much emphasis was being placed on the traditional sequences with respect to the prerequisites for the course. Were they really the necessary prerequisites? Perhaps if the matter were examined more carefully one would discover that the level of conceptualization required was such that calculus could really be taught effectively at a much earlier period. With programmed and computer assisted instruction, it may be possible to have individuals entering college with a far better background of knowledge than is now the case. In limited fields highly individualized post-doctoral instruction may be placed at the pre-doctoral level, and perhaps phases of pre-doctoral instruction can be similarly given earlier.

Student motivation is another tremendously important area. Learning aids such as programmed instruction and computer assisted instruction certainly can contribute to student motivation. Computers can be helpful not only in instruction, but also in assisting research workers in investigating various forms of presentation, various strategies in instruction, and various ways of keeping the student interested. The way these factors vary in different stages in learning should also be studied because the assumption that whatever works at one stage is going to work the same way in later stages in learning does not invariably hold.

The problem of presenting the material in a way to increase motivation and to speed up learning lends itself to these forms of assistance. We have instances in which a modification by the teacher in the way the problem is presented has resulted in more rapid learning by the student. For example, Norman Meier reported informally that the majority of sophomores who were subjects in one study came up with the wrong answer when the problem they were asked to solve was presented as follows: "A man bought a horse for $60 and sold it for $70. He bought it back for $80 and sold it for $90. What was his profit?" The students came up with the answer that his profit was $10. They said that when the man bought the horse for $60 and sold it for $70, he made $10; when he bought it back for $80, that wiped out his gain; but he sold it for $90 so he finished with a net of $10.

Then Meier presented the problem in this way: "A man bought a white horse for $60 and sold it for $70. He bought a black horse for $80 and sold it for $90. What was his profit?" Everybody said $20. Some students have the idea that there are teachers who have a knack of presenting a problem in a way that suggests the wrong answer. This makes problem solving difficult for them.

One individual stated that he had been teaching in universities for fifteen years and he had never given a lecture. He hoped he never would. He thought it was a waste of time which merely satisfied the lecturer. He believed in class discussion and would endeavor to elicit good answers to questions. One student gave him a clipping that momentarily brought him up short. It referred to a school where the teacher was using what was classified in the article as the Socratic method. The teacher wanted to bring out information from the child about Columbus discovering America. The teacher asked many questions, would not give the answer, but kept asking questions again and again. Finally, when he could not get the answer from the child, he said "What does 1492 suggest to you?" And the child said "Marked down from $15.00." The conclusion "Use the Socratic method and you will have high school students graduating from high school in six years," does not follow of course, and may merely reflect a different bias.

Occasionally the view is advanced that the program of training which often characterizes the year or two of postdoctoral education in a special area should serve as an excellent model for predoctoral education in the student's major. Reference is again to the close apprentice type relationship to the supervisor, in which education is virtually tutorial with a program that provides for individual assignments, individual rates of progress, and formal and informal consultations as needed. If this model were applied to the four year graduate program in psychology leading to the Ph.D. degree, it would presumably involve few formal academic courses, probably no core program of courses of the lecture-discussion variety at any year, but would involve research seminars, and research experience in the laboratory in an apprentice relationship with member(s) of the faculty in the area of the student's research and teaching and/or other professional interests.

It would seem that such a program would be admirably suited to provide thorough "specialization" at the graduate level but the "reasonable breadth" expected at the level of competence represented by the Ph.D. degree might become a matter of some concern and the limitation on the student-staff ratio would indeed be serious.

The quality of education found in excellent individual instruction need not suffer in new developments in educational technology. Important factors in programmed instruction and in computer assisted instruction are that the roots are deep in psychological learning theory and attention is focused on the processes by which individuals learn, rather than on the way a teacher administers a classroom.

PROJECT ELECTIVE PARTICIPATION: A PRELIMINARY REPORT

King M. Wientge

Rationale

This project, which was formally initiated in the fall semester of 1966, investigates the relationship between a variable, tentatively called "elective class participation," and academic achievement in credit courses for adults. The project is scheduled to run through the academic year 1967-68 so the findings presented in this paper are not complete and constitute a preliminary report.

A distinguishing feature between full-time undergraduates and part-time adult students seems to be "maturity," which involves not only years of chronological age but also the assumption and management of responsibilities in the family, at work, and in the community. Findings from an earlier project in adult education[1] indicated that a number of age-related variables, reflecting various aspects of "maturity," are positively and significantly related to adult academic success.

From one point of view this finding is somewhat surprising, since the older adults, with a wide variety of responsibilities, might be considered as having less time for study and class attendance than an individual with fewer commitments. On the other hand, the adult student who is accustomed to being a parent and to playing an important role in business and civic affairs seems to carry into his studies a realization of their importance and a desire to achieve. Of course, we do not know whether outside responsibilities make the adult a more serious student or whether the ability to handle credit courses is just one more manifestation of "maturity." In this project the aim is to explore a method of instruction developed to capitalize on the greater maturity of adult students.

In general, the objective of the project is to develop and evaluate a framework for adult instruction which will encourage students to be active learners in a campus environment, and take increased responsibility for their achievement in a defined subject matter area.

---

[1] Wientge, King M. and DuBois, Philip H. Factors Associated with the Achievement of Adult Students, University College Research Publication No. 5, Washington University, St. Louis, Missouri, 1964.

Specifically, the present study attempts to investigate whether adults in credit type instruction learn better under conventional or permissive conditions.

Under conventional conditions the instructor sets the stage,and full participation in a program of class activities is explicitly or implicitly demanded such as is typical of the traditional lecture-discussion method.

Under permissive conditions the instructor still sets the stage but the students may choose the degree of participation in class activities established by the instructor.

Development of Class Activities

Activities to be carried on in each weekly class under experimental conditions were produced by the assigned instructors with the support of the research staff. The activities scheduled can be categorized under the following general headings: films, measurement devices, feedback quizzes, review and lecture-discussion. Whatever instructional device or method seemed best suited for the weekly topic to be covered was utilized. The development of program content that is meaningful and attractive to adults of "maturity" is a vital basic assumption in permissive learning.

A brief description of each weekly activity was incorporated in a clsss schedule which was handed out in advance to each student enrolled in an experimental class.

Table 1 presents the student ratings of the permissive class sessions at the end of the semester. It is noteworthy that no activity was rated less than average, i.e., below 5. The highest average rating for the four experimental classes is 7.2. This activity featured movies and discussion on the brain and central nervous system. Generally this is not thought of as a popular subject-matter topic. Here the presentation had meaning and interest to all classes. This suggests that analyses of the activities with high ratings would be useful in developing future class activities for experimental sections.

Table 1 - Student Rating* of Activities by Class

	N	2	3	4	5	6	7	8	9	10	11	12	13	14	15	M
Class A	18	5.6	5.0	5.2	7.4	5.1	6.2	5.3	5.6	5.8	5.7	5.3	5.4	5.6	5.7	5.64
B	46	6.5	6.7	6.4	7.1	7.3	7.1	5.9	6.8	5.9	6.0	6.1	7.0	6.5	5.6	6.48
C	36	5.7	5.1	5.9	7.3	6.3	6.3	5.1	6.7	5.6	6.1	5.5	6.2	5.8	5.9	5.96
D	14	6.1	6.0	5.4	6.8	7.1	7.3	4.8	6.6	5.3	6.4	6.0	6.2	5.4	5.7	6.08
M	114	6.0	5.7	5.7	7.2	6.5	6.7	5.3	6.4	5.6	6.0	5.7	6.2	5.8	5.7	6.04

Activity

Week No. 2 History & development of psychology. Test battery & relation to individual differences.
3 Efficient study procedures & reading comprehension test.
4 Interpretation test battery scores; uses related to understanding indiv. differences.
5 Movies and discussion on brain and central nervous system.
6 Feedback quiz, scoring, discussion & review first 3 chapters Munn.
7 Film on heredity.
8 Administration of unique intelligence measuring instrument and its application.
9 Film on Skinnerian behavior theory.
10 History & development of concept of intelligence within psychology.
11 Lecture and discussion of learning process.
12 DuBois-Bunch Learning Test with discussion, and lecture-discussion of Forgetting.
13 Feedback quiz and review.
14 Contin. study of Forgetting w/demonstration of proactive & retroactive inhibitions.
15 Lecture and discussion on Thinking and review of course.

* Ratings are on the stanine scale, from 9 high to 1 low.
5 would represent an average rating.

Student Rating of Experimental Course

At the end of the semester all students enrolled in experimental sections were given the opportunity to: a) write comments about the course; b) express a preference whether they would again enroll in a class conducted similarly; or c) express preference for the traditional lecture-discussion method for classroom learning.

a) Comments (voluntary)

Number of students volunteering comments		69	
Nature of comments:	Favorable	36	52%
	Adequate	16	23%
	Unfavorable	17	25%

b) Would enroll if similarly conducted class offered:

Yes 43 - 36% No 27 - 22% With changes 51 - 42%

c) Prefer traditional lecture-discussion method:

Yes 67 - 74% No 24 - 26%

The comments are not entirely consistent, but seem to indicate a continuing interest in the traditional lecture-discussion method, although with quite positive acceptance of permissive instruction if certain changes are made. It remains for another year's experimentation with classes under permissive and conventional conditions to obtain data which will permit a more complete evaluation of the two conditions. This will shed more light on the interesting but conflicting appraisal afforded by this preliminary look.

Statistical Analysis

One of the most important criteria for evaluating the success of the experimental condition is the comparison between gain in learning under permissive instruction with that amount gained under conventional instruction. The data for this analysis will not be available until the conclusion of the academic year 1967-68 as indicated in the following chart. Then all classes in Psychology 9-205 taught under permissive conditions in 1966-67 can be compared with all classes in Psychology 9-205 taught under conventional conditions in 1967-68. The reverse is true for the second semester of General Psychology, 9-206.

	1966-67	1967-68
General Psychology First Semester 9-205	All classes experimental (Permissive)	All classes control (Lecture-discussion)
General Psychology Second Semester 9-206	All classes control (Lecture-discussion)	All classes experimental (Permissive)

In order to evaluate level of knowledge of psychology, 100-item objective tests covering the course content in Psychology 9-205 and Psychology 9-206 are administered at the beginning of each semester to all experimental and control classes. At the end of the semester equivalent 100-item objective tests are administered as final examinations. By applying the method of "residual gain" statistical analysis to the beginning and end of course scores a measure of gain in learning is obtained which compensates for differential prior knowledge of students and permits a more realistic comparison of knowledge gained in experimental and control classes.

In summary, this preliminary report has attempted to provide some insight into the nature of the problem, the rationale and design of the study and a look at preliminary findings which point the way toward a more comprehensive investigation for the complete research program to be completed by the fall of 1968.

LEARNING OPERATIONAL EQUIPMENT AS A CRITERION IN TRAINING RESEARCH

G. Douglas Mayo and Alexander A. Longo

It has been said that training and education differ, not so much in terms of methods and techniques as in the specificity of their objectives. This statement, we believe, can be extended to research pertaining to education and to training, as well. While personnel engaged in training research would prefer that their findings be generalizable over a wide range of educational and training situations, they do not consider this to be an essential condition. The study reported here is a case in point, and in fact capitalizes upon a condition that is essentially peculiar to naval aviation technical training. It does have some features, however, that are interesting to us, and perhaps will be to others. Primary among these features are the implementation of the concept of ability to learn operational equipment as a criterion, and the construction of a criterion measure based upon terminal performance specifications in a representative operational equipment course. These two topics will be described in greater detail in a subsequent part of the paper.

This is the third in a series of four reports describing a study which compared the performance of students receiving training of different lengths for the Navy rating or occupation of Aviation Electronics Technician R (Radar and Radar Navigation Equipment). The four reports are based upon measures of performance taken at four points in the training and occupational "pipeline." The first measure was taken at the end of the entry course in aviation electronics fundamentals. The second measure was taken at the end of a follow-on course in which radar equipment was used as a means of teaching principles and theory of radar.

The present paper is based upon measures taken in a course pertaining to maintenance of a specific piece of equipment used in operating squadrons to which the personnel receiving the course were assigned. The final report will provide information concerning the performance in operating squadrons of personnel receiving different amounts of training.

One group received the original aviation electronics fundamentals course, which was 19 weeks in length, and then received the original radar course, 11 weeks in length, for a total of 30 weeks of training before being assigned to a squadron and receiving a brief course in the specific equipment used by the squadron. The other group received a revised and shortened course, both in aviation electronics fundamentals and in radar training. The revised fundamentals course was 14 weeks in length and the revised radar course, 8 weeks in length, for a total of 22 weeks. Thus the students assigned to the revised courses received 8 weeks less training than did the students assigned to the original courses.

It is desirable for personnel to spend as small a proportion of their time in training status as possible, consistent with their acquiring the knowledge and skill necessary to performing adequately on the job. In general, time spent in training, not only is not immediately productive but requires diversion of material resources and support personnel from productive tasks. This matter has been accentuated in recent years by increased requirements for training, stemming from more complex equipment on the one hand and low reenlistment rate of first enlistment personnel on the other. Thus, it is quite important that personnel be adequately trained and at the same time that training of first enlistment personnel be held to a minimum, consistent with satisfactory performance.

The basic training plan which has emerged in naval aviation technical training retains the economy inherent in mass production for general, theoretical training, but recognizes that formalized training must be given on the specific equipments and systems the man will be expected to maintain in order for him to perform satisfactorily. Thus, following general, theoretical training in the basic schools, personnel are assigned to operating squadrons but, as a rule, are trained on the equipments used by these squadrons before reporting to the squadron. This training is conducted at the naval air station at which the operating squadron is stationed when ashore by Combat Readiness Air Groups, which in turn delegate most of the responsibility for formalized maintenance training of enlisted personnel to the Naval Air Maintenance Training Detachment located at the same naval air station.

The Combat Readiness Air Group has the same type of operational aircraft as the squadrons for which they train, but most of the training of maintenance personnel is conducted on training panels which are developed by the manufacturer of the aircraft, complete with all training materials required.

These training panels are essentially the same as the corresponding equipment or system in the aircraft but display the equipment in such a way that an understanding of the system and maintenance procedures pertaining to it are more readily acquired than would be the case if the actual aircraft were used.

Since the primary function of general, theoretical training in the basic or Class A schools is to train to the point that the graduate of the school can readily assimilate training on the operational equipment and systems he will be expected to maintain, a measure of ability to acquire information and skills pertaining to such operational equipment becomes a logical criterion in the evaluation of basic school training.

When viewed in this manner, the criterion of performance in operational equipment courses has certain advantages, as follows:

(1) A substantial period of behavior observation is provided, longer than ordinarily would be possible in a performance testing situation.

(2) The behavior observed is directly related to the objectives of the general, theoretical training and to the work the man will be doing on the job.

(3) The observation of behavior occurs before further training and experience have an opportunity to exert a differential effect upon the various individuals in the group.

If performance in operational equipment courses is to serve as an adequate measure of the effectiveness of training previously received in the basic schools, certain conditions must be met. First, it is necessary to select an equipment course that is representative, in terms of content and difficulty, of the equipment courses to which graduates of the basic school are assigned. This is true because it is difficult to compare graduates attending a number of different courses. Following the selection of a representative equipment course, the input to the course must be controlled to ensure that the personnel assigned to the course are representatives of the treatment groups to which we wish to generalize. Thus, members of the groups which we wish to compare may be assigned to the representative equipment course on a random basis or they may be matched on the basis of a pertinent measure taken before the beginning of the course. Finally, it will usually be necessary to develop or revise the measures used to assess performance in the equipment course in order to ensure their adequacy as a criterion measure. In those instances in which the conditions just described can be met, ability to learn operational equipment and systems has considerable appeal as a criterion in training research.

The sample consisted of 58 students who graduated from the Aviation Electronics Technician R (Radar) Course at the Naval Air Technical Training Center, Memphis, Tennessee, between August and November 1965. All subjects were regular Navy personnel who had had no previous Navy experience. The mean civilian educational level of the group was approximately 12 years and means on pertinent aptitude tests, such as the Navy General Classification Test, Arithmetic Test, and Electronics Technician Selection Test, were in a 60-65 range. These scores are Navy standard scores, which in the Navy standardization group had a mean of 50 and a standard deviation of 10.

A matched group design was employed which utilized 29 pairs of subjects. The matching variable was the grade made in the Aviation Fundamentals School, a course two weeks in length, completed immediately prior to training in electronics fundamentals. The correlation between the matching variable and the primary measure of performance in the present study was .41 in the case of the group that received the original, longer course in electronics fundamentals and radar, and .24 in the case of the group that received the revised, shorter course.

The original design called for 30 pairs of subjects, but after the subjects had been assigned, one man became ill and it was necessary to eliminate this pair of subjects from the study. The two treatment groups were enrolled in the course at different times. The first group previously had received thirty weeks of training in electronics fundamentals and principles of radar equipment, as described in a previous section of this paper. Six students from each of five classes completing the Aviation Electronics Technician R (Radar) Course during August and September 1965 were assigned to a one week course in the coder group portion of the P-3A IFF/Loran System, APA-89. This course, which consisted of 36 hours of actual instruction, was conducted by the Naval Air Maintenance Training Department at the Naval Air Station, Patuxent River, Maryland.

The purpose of the course was to provide maintenance personnel with instruction on the latest maintenance techniques, modifications, operational systems, and circuit analysis of the coder group portion of the IFF/Loran System in the P-3A aircraft. Upon completion, maintenance personnel were expected to be able to diagnose, troubleshoot, maintain, and service the coder group portion of the IFF/Loran System and know the safety precautions to be observed.

Because of the nature and size of the equipment involved, the maximum number of students that could be enrolled in the course at a given time, was six. Selection for the course was made about four weeks prior to graduation from the Aviation Electronics Technician R (Radar) Course. The subjects were selected in a manner which would provide a representative range of scores on the matching variable. The members of the other treatment group, the group that received 22 weeks of electronics fundamentals and radar principles training, completed the Aviation Electronics Technician R (Radar) Course during October and November 1965. Six graduates of the course from each of five classes during this period were assigned to the specific equipment course at the Naval Air Maintenance Training Detachment at Patuxent River, Maryland, in a manner similar to that described in connection with the first group. In this instance, however, members of the graduating class were paired with a member of the first group having the same or a similar score on the matching variable.

It was not possible to match perfectly in every instance on the basis of grades made in the two weeks Aviation Fundamentals Course, as had been true in earlier phases of the study. The means and standard deviations of the two groups were quite similar, however. The means were 80.52 and 80.41 for the groups receiving 30 weeks and 22 weeks training, respectively. The corresponding standard deviations for the two groups were 6.42 and 5.83. Expressed as a correlation coefficient, the correlation between pairs on the matching variable was .95.

Immediately following completion of the Aviation Electronics Technician R (Radar) Course, the graduates were given two weeks leave, at the end of which they reported for the course at the Naval Air Maintenance Training Detachment, Patuxent River, Maryland.

Two primary characteristics were sought in the criterion measure. First, it was essential that the equipment course be representative of the equipment maintenance courses taken by graduates of the Aviation Electronics Technician R (Radar) Course in terms of content and difficulty level. Second, it was necessary that performance in the course be measured as adequately as possible. In satisfying the first of these requirements six of the best informed personnel at the headquarters of the Naval Air Maintenance Training Group were asked to nominate several courses that they considered most representative of the type of courses that were taken by graduates of the Aviation Electronics Technician R (Radar) Course. This action was taken to reduce the large number of courses offered by the Naval Air Maintenance Training Group to a manageable number.

The nominations by the panel of six judges, first performing independently and later as a panel of advisors, produced a list of ten equipment courses that were considered most representative of the courses taken by graduates of the Aviation Electronics Technician R (Radar) Course. Since assignment to equipment courses is based upon the equipment with which the graduate will be working in operating squadrons, these courses tended to pertain to the equipment most widely used within operating squadrons. This list of ten courses pertaining to specific equipments was then duplicated and each member of the advisory panel was asked independently to rank the courses in terms of their representativeness. Three courses stood out as the most representative, and a choice was made among these three on the basis of overall feasibility, which included measurement aspects, and location and length of the courses.

Measurement of performance in the course involved a procedure that had not been used previously in training research within the Naval Air Technical Training Command. The end product of the procedure, and the primary measure of performance used in the study, was a "criterion test," the items of which corresponded to the "specific-behavioral objectives" of the course. The terms criterion test and specific behavioral objectives are used with their somewhat specialized connotation derived from programmed instruction.

In developing the performance measure, the senior instructional programmer for the Naval Air Maintenance Training Group, a chief petty officer with many years of electronics experience, attended the entire course and identified the specific behavioral objectives which the course sought to achieve. He then constructed one or more "constructed response" items (as contrasted with multiple choice type items) designed to measure each of the specific behavioral objectives of the course. The objectives and test items were discussed with the personnel conducting the course and agreement was reached as to their pertinence and inclusiveness.

Following the completion of each of the ten classes, the test was scored by the chief petty officer who had constructed the test. The test contained 35 items, and the score assigned each subject was the number of items answered correctly out of the 35 items on the test. While this test was considered to be the best measure of performance in the course, two other scores also were collected. These were scores on performance tests of troubleshooting and of alignment of the equipment. These performance tests were already in use and were not altered for purposes of study.

The results of the comparisons that were made between the two groups in terms of their performance in a representative operational equipment course are shown in Table 1. A statistically reliable difference was not found between the group that previously had received the original, 30 weeks course and the group that had received the revised, 22 weeks course. In the case of the measure that was considered to be the best measure of performance in the specific equipment course, the programmed instruction type "criterion test," the mean number of items answered correctly on the 35 item test was 27.77 for the group that had taken the original course and 28.37 for the group that had taken the revised course.

TABLE 1

Comparison of the Two Matched Groups in the Operational Equipment Course (N = 29 pairs)

Previous Training Received	"Criterion Test"			Alignment			Troubleshooting		
	Mean	S. D.	t	Mean	S.D.	t	Mean	S.D.	t
Original Course (30 weeks)	27.77	3.88		74.72	3.95		73.90	4.19	
Revised Course (22 weeks)	28.37	3.13	.76	76.59	4.51	1.82	75.90	6.20	1.59

Note: A t value of 2.01 is required for significance at the .05 level; 2.67 at the .01 level.

Similar results were found in the case of the two performance measures normally used in the course. Mean scores on the alignment test were 74.72 for graduates of the original course, as compared with 76.59 for graduates of the revised course. On the troubleshooting measure the mean scores were 73.90 and 75.90 for graduates of the original and revised course respectively.

There are several points that appear to warrant comment. The first, and most obvious, is that the study gives no evidence of the original, longer course providing better preparation for the operational equipment course than the revised, shorter course provides. In fact, the means of all three measures taken upon completion of the equipment course were slightly, but not reliably, higher for graduates of the revised course than for graduates of the original course.

The concept of using performance in a representative equipment course as a measure of the adequacy of previously received basic training was applied in this study for the first time in naval aviation technical training, and as far as the investigators are aware, anywhere. On the basis of experience with the method, it appears to be feasible from the standpoint of practical application, and to provide information under conditions that are better controlled than ordinarily is possible in evaluation studies conducted in the fleet.

While the methodological aspects of the study doubtless are of primary interest to this conference, in your capacity as individual taxpayers, you may be interested to know that the project as a whole resulted in something over 20,000 man-weeks annually being available for productive performance in the fleet as opposed to this time being spent in a "consumer status" in training.

EVALUATION OF A PARTIALLY SELF-PACED COURSE

Kirk A. Johnson

It is generally felt that education or training is most effective when it is tailored to the characteristics or abilities of the individual student. The most rudimentary form of adjustment can be achieved by simple variations in training time. Even this, however, has been relatively rare in military training situations. The adjustments that are made in situations of this type have generally taken the form of night schools and repetitions for students who are having difficulty. The very gifted, highly motivated student has occasionally been allowed to study the course materials on his own. In recent years, two or three track training systems have been developed for some courses.

The most obvious obstacle to still more individualized forms of training has been the lack of efficient means for insuring control over student behavior. Many of these difficulties have been reduced, however, through recent developments in educational techniques. As a result, a second obstacle has become much more prominent. This is the difficulty in scheduling instruction in such a way that the student can be shifted without delay between portions of the curriculum that can be taught most efficiently by means of individualized instruction and other portions of the curriculum in which individualized instruction would be either prohibitively expensive or less effective than alternative means of instruction. In certain educational situations any gaps that might appear as a result of these scheduling difficulties could be filled with special supplementary instruction, but "enrichment" of this type would be of questionable value in most military training situations. As a result of these difficulties, the more individualized forms of instruction have been used in situations that preclude a realization of their potential efficiency, or, in relatively few cases, in courses in which the entire course could be taught by means of such techniques.

I would like to present a brief description of the first individually paced course to be developed within the Naval Air Technical Training Command. The course was the Airborne Radio Code Operator Course in which the student learned military communications procedures and how to send and receive International Morse Code. During the first part of the course, which lasted about one week, the student learned the symbols used in Morse Code. During the second part, which lasted about three weeks, the student practiced sending and receiving groups of random alpha-numeric symbols. During the third and final part, which lasted about half a week, the student practiced sending and receiving messages of the type that he would encounter in an operational military environment. Military Communication Procedures was taught by means of 30 lecture-discussion sessions distributed over the last two parts of the course.

There were several reasons for selecting this course for conversion to a more individual format. First, the people who have done work with the teaching of Morse Code have repeatedly recommended that courses of this type be geared to the progress of individual students. Of equal importance were the various features of this course that made such a conversion somewhat easier than it might have been otherwise. This course was the terminal course in a sequence, so there was no need to worry about phasing the students into another course; when they graduated they could be sent directly to their assignments in the field. The practice on random code groups was already being handled by means of tape recorders, so a major portion of the course could be presented on an individualized basis without further modifications. Finally, the material on Military Communications Procedures lent itself readily to teaching by means of programmed instruction.

In the individually paced version of the course, the first week of the course was taught in the conventional manner. During the next phase the students were permitted to progress at their own speed through practice on the random code groups and through a set of programmed instruction booklets that covered the material on Military Communications Procedures. During this phase the instructor's role was limited to that of a monitor. As soon as the student had reached an acceptable level of proficiency in code reception he was given his final performance test; as soon as he had completed the programmed instruction booklets he was given his final written test. If he passed these tests he was placed on orders. He graduated immediately prior to his departure for his new assignment. The interval between his passing of the tests and his graduation, which was generally about two days long, was spent in practice on sending and receiving military messages.

Evaluations in the field are generally beset by administrative problems, but in the present case these problems were more serious than most. During the preliminary evaluation of the programmed booklets it was learned that the course was being transferred from Memphis in an effort to rectify certain deficiencies in the assignment of personnel that had existed for a number of years. We were able to postpone this transfer until we could gather some data on the individually paced course, but much of our experimental control went by the board.

Data were gathered on 88 entering students in the regular course during a period prior to the completion of the programmed instruction booklets. Data were also gathered on 66 entering students during a period in which the programmed instruction booklets were being used to teach Military Communications Procedures. There was no individual pacing during this period. Finally, data were gathered on 28 entering students in the individually paced version of the course.

The three groups were evaluated in terms of attrition, class days per graduating student, receiving speed after 15 days, and scores on two special examinations. Both examinations were based on the same course profile. The short answer test consisted of items drawn from the 257 objectives covered by the programs. The multiple choice test consisted of items drawn from the pool used in the construction of the regular course examinations.

Scores on these criteria have been summarized in Table 1.

TABLE 1

Group Means on Six Criteria
(Unmatched Groups)

Criteria	Groups		
	Regular	Program	Ind. Paced
% Drops	9.1	9.1	17.9
% Setbacks	10.2	16.7	0.0
Days per Graduate	25.8	24.0	21.8
Receiving Speed	9.1	9.3	9.3
Short Answer Test (%)	77.3	83.2	82.1
Multiple Choice Test (%)	81.1	79.0	77.1

The total percentage of drops and setbacks was less in the Individually Paced group than in the other groups, but the percentage of drops, which is probably a somewhat more expensive form of attrition, was higher. These drops tended to occur quite early in the course, at a point when none of the students is very proficient at receiving code and prior to any testing on Communications Procedures. Since the decision to drop a student is based on fairly subjective considerations, these data on attrition may not provide a very reliable index as to the actual value of the various courses. Days per graduate was computed by dividing the time spent in class by all entering students by the number of graduates, so these figures reflect differences in both training time and attrition. The Conventional course required about 26 days per graduate. This figure was reduced by about 2 days in the Program course, and by 2 more days in the Individually Paced course. The two groups that used the programs were quite similar to one another on each of the remaining criteria. They were about the same as the Regular group on receiving speed, about 5 percentage points better on the short answer test, and about 3 percentage points worse on the multiple choice test.

These data were fairly encouraging but further investigations indicated that students entering the Individually Paced course tended to be somewhat better than students entering the other courses in terms of both general aptitude and performance in previous courses. In an effort to compensate for these differences, students who graduated without setbacks were matched on the basis of their average grade in the three courses preceding the course under investigation. It was possible to form 20 matched triads. The data from these students on three of the criteria have been summarized in Table 2.

TABLE 2

Group Means on Three Criteria
(Matched Groups)

Criteria	Group Paced		
	Regular	Program	Ind. Paced
Receiving Speed	9.2	9.6	9.1
Short Answer Test (%)	77.5	83.5	83.1
Multiple Choice Test (%)	82.5	78.9	78.0

These data are not too different from those found with the unmatched groups. Receiving speeds are still about equal; the differences between the groups using programs and the regular group have decreased slightly on the short answer test and increased slightly on the multiple choice test.

Although inferential statistics are not particularly meaningful in situations of this type, it was found that group differences on the short answer test had a probability that was slightly below .05 and that group differences on the multiple choice test had a probability that was slightly above .05. Differences in receiving speed were far from significant.

In summary, the Individually Paced Course led to a reduction of about 16% in the length of the Regular Course. The primary cost was a drop of about 4.5 percentage points on the multiple choice test, but even this is probably an overestimation, since for the Individually Paced students this test did not reflect the knowledge they might have picked up during their last two days of practice on military messages. The 9% reduction in the length of the Program course could be viewed as pure profit.

Unfortunately this course was developed by circumventing rather than solving most of the problems that are associated with Individually Paced courses.

THE QUIZ AS A TEACHING DEVICE

Leon Clodfelder

The quiz and feedback from the quiz have been the subjects of many investigations dating back to at least 1923. In a review of the literature in 1940, Stroud concluded that "the mere taking of an examination results in genuine learning—perhaps as much, per unit time as any other kind of activity." However, Noll (1939) was not as optimistic. He concluded that although both college students and instructors believed that quizzes improved achievement, evidence was lacking for such a belief. He did feel that there may have been a tendency for less able students to profit by taking quizzes.

Standlee and Popham (1960) gave four reasons for the quiz's positive contribution to classroom learning; 1) it increases motivation; 2) knowledge of results are given; 3) it gives structure to the student or tells him what is important; and 4) it enforces activity with respect to the material. They felt all of these conditions must be present for the quiz to be effective. In most studies, variables not under scrutiny have not been adequately controlled so that different interpretations are proposed for similar data.

Jones (1923) followed by Spitzer (1939) and Sones and Stroud (1940) felt that the quiz offered a recall technique for material already learned. They all found that by giving groups a test shortly after a lecture or the reading of an article, scores on a retest delayed up to six weeks could be increased significantly above scores for groups which took the test for the first time. One cannot tell from these studies whether the quiz per se or practice on the material inherent in the quiz is responsible for the results. The quizzed groups, obviously, had more practice since they had taken the criterion test at least one more time than the non-quizzed groups. However, in the Sones and Stroud study some of the non-quizzed groups reread the article in lieu of taking the quiz.

Michael and Maccoby (1953) reasoned that if practice were the important variable only items on the criterion test repeated from the quizzes which were introduced during breaks in a film would be affected. If motivation were increased by taking a quiz, both repeated items and new items would be affected. Because quizzed groups did better than controls only on the repeated items, the authors concluded that the quiz increases learning because of the practice it entails on specific items.

Rothkopf's (1965) approach to the experimental task was similar to that of Michael and Maccoby, except that he was interested in what he called mathemagenic behaviors. These are behaviors in which the learner supposedly engages while he is learning, and includes such things as conceptualization of the material and mnemotechnic devices. He compares these with the effective stimulus in psychophysics. Rothkopf reasoned that such activity would be strengthened by the introduction of a short quiz during or shortly after reading the material, thus enhancing learning not only of the specific items of the quiz but also of new items on the criterion test.

To test his hypothesis Rothkopf inserted a two-item quiz either before or after each of seven sections of a chapter which subjects read. Two criterion tests were given, one consisting of questions not used in the quizzes and another of items repeated from the quizzes. Finding that only the group with the quiz shortly after each section did better on new items as compared with control groups, he concluded that general facilitative skills are learned over and above practice on specific content. These skills or mathemagenic responses were believed to be elicited and strengthened by the quiz given shortly after the reading task.

Both Rothkopf, and Michael and Maccoby reported that immediate knowledge of results on the quiz affected only the score on the repeated items and this effect was additive to the effects of the quiz. Support for mediational responses by subjects akin to Rothkopf's mathemagenic behavior is given by some studies on delay of knowledge of results. Brackbill, in a series of verbal discrimination studies with elementary school children [Brackbill, Bravos, and Starr (1962); Brackbill, Isaacs, and Smelkinson (1962); Brackbill and Kappy (1962); Brackbill, Wagner, and Wilson (1964)] consistently found that short delays (8 seconds) in knowledge of results during acquisition enhanced retention as measured by trials to relearn. In the study more related to school curricula, Brackbill, et al, (1964), using French-English pairings, found that delay enhanced the retention of the most difficult items relatively more than it did the easier ones.

Brackbill interpreted her results as indicating that subjects capable of response-produced cues retain more if these cues are utilized during acquisition. Delay of knowledge of results increases the opportunity for subjects to utilize these mediational processes. Further, the more mediational processes needed for retention, the more delay may help.

English and Kinzer (1966) sought to test the results of the Brackbill studies using meaningful verbal materials and college students as subjects. They had their subjects read different articles and immediately quizzed them, administering knowledge of results immediately, delayed one hour, delayed two days, or delayed one week. A post-test was given two weeks after knowledge of results. More gains were made from quiz to post-test by those groups whose knowledge of results was delayed one hour or two days for three of four articles. The fact that one article produced no results suggests a material x delay interaction such as Brackbill et al had found.

English and Kinzer did not vary the placement of the quiz and dealt only with repeated items. If the mediational responses of Brackbill and the mathemagenic behaviors of Rothkopf are the same or at least overlapping, one would expect that similar results could have been produced by varying the delay of the quiz from the reading of the article. In addition it is surmised that the effects would have included new items over the material in addition to repeated items. The latter situation is probably most prominent in the classroom situation. It is the purpose of the present study to see if the results obtained with reading materials can be applied to the lecture-discussion classroom and to explore the effects of delaying the quiz from the time of original learning. Further, an attempt is made to assess interaction between delay of quiz and delay of knowledge of results.

A short four to six item quiz was given either immediately after the conclusion of each lesson in physics or delayed one day. Knowledge of results in the form of correct answers was given immediately after the quiz or delayed one day. In addition, feedback in the form of number of correct answers was given one day after the quiz to one group under each quiz condition. Thus, the quiz was varied two ways and knowledge of results was varied three ways setting up a 2 x 3 factorial design utilizing six groups.

Each treatment group was one section of class 708 of the AMFU(A) school at Memphis Air Station. Subjects were 114 sailors and marines enrolled in this class and assigned alphabetically to the different sections. The school consists of four subjects integrated throughout its four weeks duration. Only the lessons in physics were involved in the study, which totaled eight. Each lesson consisted of two to four fifty-minute lecture periods, depending on the lesson.

Treatments for the six sections were as follows:

Quiz	Section		
Immediate	A	B	C
Delayed	D	E	H
Knowledge of Results			
None	A	D	
Immediate	B	E	
Delayed	C	H	

A sixteen-item pretest was administered the week previous to the subjects' commencing class in AMFUA(A) school. During the second week of class, with the commencement of the lessons in physics, the differential treatments were begun. Subjects were given five minutes to take each quiz given either directly following the lesson or the next day. Immediate feedback (correct answers) was given by the subjects' opening the left margin of the test paper on which the correct answer was given for each question and which had been folded back. They then could study the questions for five minutes. This left margin had been cut off under the delayed feedback conditions. To provide delayed feedback the left margin with the correct answers was stapled to the student's paper and returned the next day. Again five minutes were allowed for the student to study the questions and correct answers. Subjects who received only the number correct on each quiz were handed slips of paper the next day bearing their name, quiz number, and number of items correct.

Criterion was a 46 item final given the last day of class. No items from the quizzes were repeated on this test. It included the regular course final of fifteen items, plus the sixteen items of the pre-test, plus fifteen items which were added from the school's item bank. In fact, all items used were items drawn from the school's item pool with the exception of a few on the quizzes which were constructed by the experimenter from the detailed lesson guide.

If the quiz is effective in and of itself to strengthen mathemagenic behaviors and the latter's elicitation depends upon the opportunity subjects have between original learning and the introduction of the quiz to engage in such, then it is to be expected that a delayed quiz during acquisition would produce higher final scores than would an immediate quiz. Since specific quiz items are not repeated knowledge of results should have no effect except when it is not redundant to the quiz. Thus an immediate quiz with delayed knowledge of results would be expected to approach the delayed quiz conditions.

An analysis of variance was computed on the pre-test scores to ascertain if there were any differences between sections in initial level. This analysis revealed no significant differences among groups. Also, Bartlett's test for homogeneity of variance was applied for the final and it showed no significant heterogeneity of variance among groups.

The analysis of variance on the final revealed significant differences beyond the .01 level between the two quiz conditions. No significant differences were found between the three KR (knowledge of results) conditions nor was there any quiz x KR interaction. To assess just where these differences lay, Duncan's new multiple range test was applied. It revealed no significant differences at the .05 level among the three delayed quiz groups or among the three immediate quiz groups. Superiority of each delayed quiz condition over each immediate quiz condition was significant.

Such results follow predictions with the exception that the immediate quiz/delayed KR group was predicted to do as well as the delayed quiz groups and significantly better than the other immediate quiz groups. No such interaction seemed to exist.

The above results are only tentative. A replication is now in progress with the sections receiving the immediate and delayed quizzes reversed from that here reported. Consequently, no definite conclusions or interpretations have been reached. It may be that these results, if they hold up in the replication, can be explained in terms of bodies of knowledge on which there already are a large amount of data such as distributed practice. Since mathemagenic behavior seems to include many things, it would be wise to explore its relationship with these areas anyway. If indeed the results are duplicated, it would lend support to the notion that the quiz serves a similar function with lecture materials as it does with reading materials, providing it is introduced close to the original presentation of the material.

TABLE 1

Means and Standard Deviations of Final Examination

Section	A	B	C	D	E	H
Condition	IN	II	ID	DN	DI	DD
Mean	31.48	32.00	31.67	35.00	36.06	36.05
S. D.	4.89	6.15	5.53	5.13	4.53	4.05

TABLE 2

Results of Analysis of Variance on Final Examination

Source	SS^2	df	M^2	F
Quiz	457.41	1	457.41	16.85*
Feedback	15.99	2	8.00	--
Q x F	0.85	2	0.43	--
Within	2931.41	108	27.14	

*Significant at .01 level

TABLE 3

Results of Duncan's New Multiple Range Test at .05 Level

Section	A	B	C	D	H	E
Condition	IN	II	ID	DN	DD	DI

Means connected by same line are not significant.

REFERENCES

Brackbill, Yvonne, Brovos, A., and Starr, R. H. Delay-improved retention of a difficult task. *Journal of Comparative and Physiological Psychology*, 1962, 55, 947-952.

Brackbill, Yvonne, Isaacs, R.B., and Smelkinson, M. Delay of reinforcement and the retention of unfamiliar meaningless material. *Psychological Reports*, 1962, 11, 553-554.

Brackbill, Yvonne and Kappy, M.S. Delay of reinforcement and retention. *Journal of Comparative and Physiological Psychology*, 1962, 55, 14-18.

Brackbill, Yvonne, Wagner, J.E. and Wilson, Dorothy. Feedback delay and the teaching machine. *Psychology in the Schools*, 1964, 1, 148-156.

English, R. and Kinzer, J.R. The effect of immediate and delayed feedback on the retention of subject matter. *Psychology in the Schools*, 1966, 3, 143-147.

Jones, H.E. Experimental studies in college teaching. Archives of Psychology, 1923, 10(68).

Michael, D. and Maccoby, N. Factors of influencing verbal learning from films under varying conditions of audience participation. *Journal of Experimental Psychology*, 1953, 46, 411-418.

Noll, V.H. The effects of written tests upon achievement: an experiment and a summary of evidence. *Journal of Educational Research*, 1939, 32, 345-358.

Rothkopf, E.Z. Some theoretical and experimental approaches to problems in written instruction. In J. D. Krumboltz (ed.) *Learning and the Educational Process*. Chicago: Rand McNally and Co., 1965, pp. 193-221.

Sones, A.M. and Stroud, J.B. Review with special reference to temporal position. *Journal of Educational Psychology*, 1940, 31, 665-676.

Spitzer, H.F. Studies in retention. *Journal of Educational Psychology*, 1939, 30, 641-656.

Standlee, L.S. and Popham, W.S. Quizzes' contribution to learning. *Journal of Educational Psychology*, 1960, 51, 322-325.

A RESIDUAL GAIN MEASURE OF REMINISCENCE

Edward N. Peters

In the last several years a number of articles have appeared in the Journal of Educational Psychology reporting studies of the relationship between reminiscence and classroom learning. These articles have all reported the finding of significant relationships between reminiscence on a motor learning task and classroom learning. The investigators were expecting such relationships on the basis of some of the theoretical work of Eysenck and his co-workers. Their work has also lead to the investigation of the relationships between reminiscence and certain personality variables, an area of research in which significant relationships have been found by a number of investigators.

Two different methods have been used to explore the relationships between reminiscence and outside variables. One of these is correlational using a crude gain measure of the improvement over the rest period during which the reminiscence occurs. The crude gain measure consists of a measure of performance on a post-rest trial minus a measure of performance on a pre-rest trial. The manner in which such crude gain measures can lead to spurious correlations with outside variables has been previously documented by DuBois and Manning. Such correlations are a function of the variances of the pre and post-rest measures and their correlations with the outside variable. Thus such correlations as have been found by previous investigators between reminiscence and outside variables could be the result of a correlation between performance on the task and the outside variables (classroom learning or personality) and a change in the metric on the task at different levels of performance.

The other method used to investigate relationship between reminiscence and classroom learning has been analysis of variance. In this method, two groups, selected for good and poor classroom learning in some subject, have been compared using a repeated measures analysis of variance on the pre and post-rest trials of a motor learning task. This method has also yielded significant results in the form of significant interactions between learning groups and trials over rest. These results have been interpreted as support for a relationship between reminiscence and classroom learning. However, a correlation between classroom learning and performance on the motor task, plus a change in the metric over rest could also yield these results.

The appropriate method of studying correlates of change in learning situations with other variables is the residual gain technique. The questions which the application of this technique might help answer in this context are:

1. Are there individual differences in reminiscence independent of differences in task performance. In other words, is there any variance in post-rest measure on a motor learning task which is not predictable from pre-rest performance measures?

2. If there is such variance how is it related to classroom learning among naval recruits at Navy, Memphis?

To help provide answers to these and other questions, 277 enlisted men enrolled in classes at Navy, Memphis took a motor learning task, the inverted alphabet printing task, with two ten minute periods of rest inserted to provide two separate measures of reminiscence. The two residual gain measures of reminiscence over these rest periods correlated .16, significant at the .01 level, when the first post-rest trial residualized with respect to the last pre-rest trial as the measure of reminiscence. However, when the last post-rest trial was residualized with respect to the average of the last four pre-rest trials, the two measures of reminiscence only correlated .07 with each other. This latter correlation does not even approach significance. Thus the null hypothesis, that there are no individual differences in reminiscence which are not predictable from pre-rest performance, cannot be rejected.

To further explore the situation a factor analysis of the scores on eighteen trials of the motor task was carried out. The correlations among the trials form a nearly single factor matrix in which the first factor removed by principle axis factor analysis is a factor of performance on the task. The residual correlation between the two post rest trials when this factor is removed is -.02. This suggests that if there is any variance in the post-rest trials that is not general performance, it is not the same in these two post-rest trials.

Conclusions

1. Residual gain is the appropriate method for studying the relationship of reminiscence to outside variables, as opposed to the methods which have been used. However, the post-rest trial should be residualized with respect to several of the pre-rest trials if one wishes to remove pre-rest performance variance from the post-rest variance.

2. Those interested in studying the relationship of reminiscence to classroom learning or other variables should first demonstrate that there are individual differences in reminiscence independent of pre-rest performance. This study failed to find such differences on a commonly used group reminiscence task.

A STUDY OF RETENTION OF PROGRAMMED LEARNING

William C. Orr

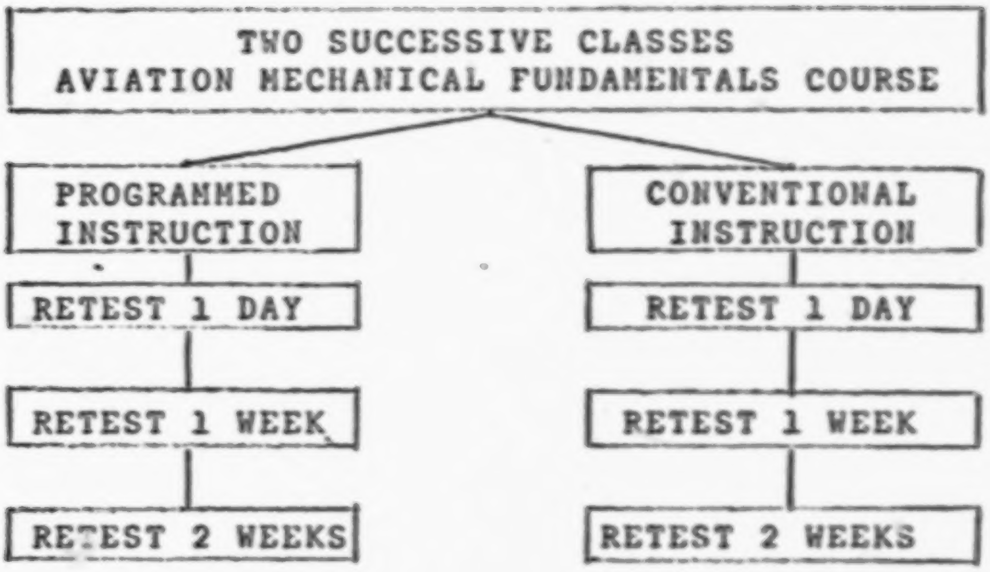
A little over a year ago at this time, I was reading some of the literature on programmed instruction, and it occurred to me that a significant variable has been overlooked in the research on this relatively new instructional method. This was the variable of retention. It would seem that if programmed instruction is considered useful, the retention of programmed learning should compare favorably with that of material learned by the conventional lecture method. This study was designed to investigate this idea.

The experimental design is shown in Figure 1. As can be seen from this, two successive classes of the Aviation Mechanical Fundamentals course were used as subjects. Each of the classes was divided into a programmed group and a conventional lecture group. Each of these groups was, in turn, divided into three retest groups. One group was retested at the end of one day, another group at the end of one week, and the last group at the end of two weeks. The subjects in the retest groups were matched on their final average from the Aviation Fundamentals course. This is a two week course in basic aviation fundamentals which all students must take before going on to their specialized courses. Studies by the Staff Research Unit at the Naval Air Technical Training Center at Memphis have shown this variable to be the best predictor of success in the specialized courses. The three separate retest groups were used in order to avoid the practice effect resulting from repeated retesting with the same test. The program used was Atomic Structure and Static Electricity. It is a linear program with a suggested reading time of 106 minutes. The test used was a 32-item 4-choice test.

Figure 2 outlines the testing and scoring procedures. A pre-test was given two weeks prior to instruction to assess the subjects' pre-instruction knowledge of the material to be covered. The post-test was administered immediately upon completion of the instruction. This was handled somewhat differently in the two groups. The program subjects were instructed to raise their hand upon completion of the programmed booklet, and they were given the test. Since much of the repetition and organization involved in learning is accomplished for the student in the programmed presentation, it was felt that the conventional subjects should also be given an opportunity to do this. A fifteen minute study period was, therefore, given to the conventional group. This served further to equate the two groups. The retests were administered as just described.

Figure 1

EXPERIMENTAL DESIGN

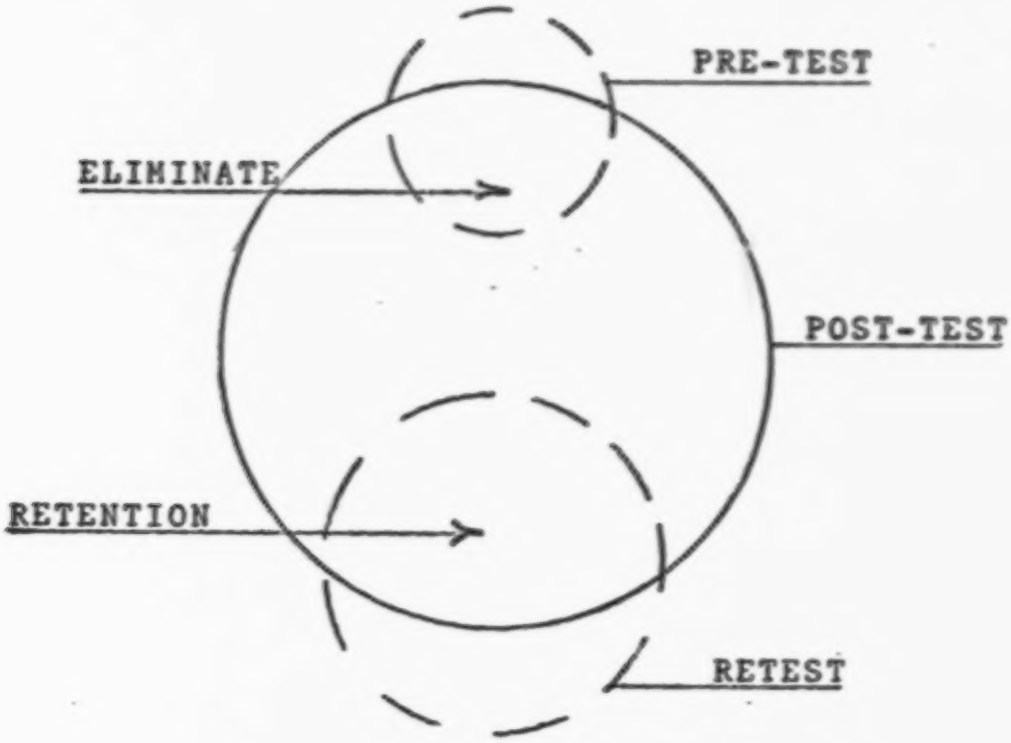

TWO SUCCESSIVE CLASSES
AVIATION MECHANICAL FUNDAMENTALS COURSE
PROGRAMMED INSTRUCTION
CONVENTIONAL INSTRUCTION
RETEST 1 DAY
RETEST 1 DAY
RETEST 1 WEEK
RETEST 1 WEEK
RETEST 2 WEEKS
RETEST 2 WEEKS

SUBJECTS IN RETEST GROUPS MATCHED USING THEIR FINAL AVERAGE FROM THE AVIATION FUNDAMENTALS COURSE

Figure 2

TESTING AND SCORING PROCEDURES

I. Pre-Test (Two weeks prior to instruction)
II. Post-Test (Immediately upon completion of instruction)
III. Retest (1 day, 1 week, 2 weeks)


PRE-TEST
ELIMINATE
POST-TEST
RETENTION
RETEST

The retention score is shown by means of Venn diagrams in Figure 2 and was derived as follows:

1. Items correct on pre-test and correct on post-test were eliminated because these are items which, it is assumed, the subject knew prior to instruction. The remaining items correct constitute the post-test score.

2. Items correct on the post-test and correct on the retest constitute the retention score.

As is clearly seen from Table 1, there were no significant differences at any of the retest intervals or between the two groups as a whole. A few cautions should be exercised in interpreting this data. First, it may be that two weeks is not a long enough retest interval for any differences to materialize. In this study, a two week retest interval was the maximum due to graduation from the course and subsequent assignment to the fleet. A second factor to consider is the type of instruction. Obviously, the material used here is amenable to a programmed presentation. Material of a different kind may not be learned as well through programmed instruction, and hence would yield quite different results.

In view of the results of studies which have shown programmed instruction to be a time saving teaching method, the results of this study seem further to support the idea that programmed instruction should be considered as an effective and valuable teaching method.

Table 1 - Results

	Conventional			Program		
	M	SD	N	M	SD	N
Items correct on post-test and correct on retest						
1 day retest	24.7	3.5	35	25.9	3.2	35
1 week retest	25.3	2.7	29	24.1	3.9	29
2 week retest	24.4	3.8	25	25.0	2.5	25
Total (all three groups combined)	24.8	4.1	89	25.1	4.3	89

References

Little, E. B. Pre-test and retest scores in retention calculation. *Journal of Experimental Education*, 19, 1960, 161-167.

Mayo, G. D., and Longo, A. A. The comparison of conventional and programmed instruction in teaching elements of electrical physics. Naval Air Technical Training Command, *Technical Bulletin*, 1964.

Prediction of Class A school success with "P" school grades. Naval Air Technical Training Command, *Technical Bulletin*, 1952.

CONSUMER ACCEPTANCE OF PROGRAMMED INSTRUCTION

Howard L. Fleischman

Along with determining the effectiveness of programmed instruction and investigating new methods and applications, it may be desirable to assess the attitudes of consumers when one is interested in making the most effective use of programmed instruction. Since programmed instruction is a vital part of the training effort in the Naval Air Technical Training Command, a good opportunity for the determination of consumer acceptance was available.

A questionnaire was assembled consisting of 28 questions and statements pertaining to programmed instruction. Subjects were asked to respond to the questions and statements from a list of five alternatives which ranged from the very favorable to the very unfavorable. In order to obtain answers which were as honest as possible, the subjects were told that the results would be used for research purposes only and that they should not put their names on the answer sheets as this was to be an anonymous questionnaire.

The subjects consisted of three types of consumers: students, instructors, and training administrators. The students were Navy and Marine enlisted men, for the most part directly out of recruit training, who were attending the aviation technical training schools of the command. The instructors were Navy and Marine non-commissioned officers. The training administrators were either commissioned officers or civilian education specialists.

The participants in this study came from eight schools of the Command. These schools were:

1. Aviation Electronics Fundamentals School at Memphis, Tennessee

2. Mechanical Fundamentals School also at Memphis

3. Air Controlman's School at Glynco, Georgia

4. Aviation Ordnance School at Jacksonville, Florida

5. Aerographer's School at Lakehurst, New Jersey

6. Aviation Boatswain's Mate School also at Lakehurst

7. Aviation Electrician's School at Jacksonville, Florida

8. Photographer's School at Pensacola, Florida

The number of hours or programmed instruction in these courses ranged from 55 in Aviation Electronics Fundamentals School to approximately 25 in some of the others. The same questions werè presented to all three types of consumers except for a few where the wording had to be changed to make the item more appropriate to the particular type of consumer.

The subjects responded to the items on a scale that ranged from favorable to unfavorable opinions of programmed instruction. To score the questionnaire, the alternative responses were credited with 4, 3, 2, 1 or 0 points from the favorable to the unfavorable end, respectively. A neutral or uncommitted response to a question was scored as 2 points. The sum of the item credits for the 28 questions represents the subject's total score.

Results and Discussion

Table 1 shows the results obtained from the questionnaire for each sub-group. The maximum or most favorable score attainable is 112, with 56 being the neutral response. Taking a look at the individual means, the student groups scored significantly above the neutral point in each of the schools with the Aviation Electronics Fundamentals School yielding the highest mean (80.29).

TABLE 1
SUMMARY DATA BY GROUPS

	Schools							
	Av. Fund.	Mech. Fund.	Air Control.	Av. Ord.	Aerog.	Av. Boatswn.	Av. Elect.	Photo.
Students								
$\bar{X}$	80.29	76.35	67.50	68.69	62.56	77.21	68.43	67.89
SD	13.11	15.49	18.55	19.26	19.85	16.70	17.32	18.79
N	128	167	126	118	111	98	165	120
p*	.01	.01	.01	.01	.01	.01	.01	.01
Instructors								
$\bar{X}$	49.94	54.33	54.28	68.96	64.66	80.57	67.31	51.55
SD	18.36	19.13	16.19	15.52	22.93	10.15	15.23	21.21
N	51	51	14	20	3	7	60	18
p	.05	NS	NS	.01	NS	.01	.01	NS
Administrators								
$\bar{X}$	51.84	60.40	67.14	62.33	62.50	70.60	69.11	43.25
SD	18.14	20.98	10.49	18.22	18.49	12.17	23.32	6.25
N	19	10	7	9	6	5	9	4
p	NS	NS	.05	NS	NS	.05	NS	.05

*Probability of significant difference from neutral response.

Three of the instructor groups scored significantly above the neutral point; four had mean scores not significantly different from the neutral point; and one instructor group (the one from Aviation Electronics Fundamentals School) scored significantly below the mean. It is interesting to note that the school where the instructors scored significantly below the mean was the same school whose students gave the highest approval to programmed instruction.

Going on to the training administrator groups, we find that two groups scored significantly above the neutral response; five administrator groups did not show a significant difference; and one group of administrators (the Photographer's School group) scored significantly below the neutral point.

To test for differences between the groups, an analysis of variance for groups with unequal frequencies was computed. The summary table for this analysis is shown in Table 2. Differences were found to exist between types of consumers at the .01 level of significance and also between schools at the .01 level of significance. No interaction effect was found.

TABLE 2

ANALYSIS OF VARIANCE SUMMARY TABLE

Source of Variance	SS	df	MS	F	p
Type of Consumer	6,642.24	2	3,321.12	10.72	.01
School	10,567.20	7	1,509.60	4.87	.01
Interaction	820.46	14	58.60		NS
Within	403,068.00	1302	309.57		
Total	421,097.90	1325	317.80		

In summarizing the results obtained from the questionnaire, we can say that the students look upon programmed instruction very favorably. This was true of all of the schools that were sampled. Instructors and administrators were found to be less favorable. However, in only one school each did the instructors and administrators have a significant negative reaction towards programmed instruction.

What I would like to do now is to present nine individual questions out of the 28 in the whole survey which I feel might be of interest. The first question reads:

"In comparing work done using programmed instruction with studying in regular textbooks, I feel that with the same amount of time and effort I learned:

1. Much more with programs.
2. Somewhat more with programs.
3. About the same.
4. Somewhat more from textbooks.
5. Much more from textbooks."

The results showed that all groups averaged above the neutral point, that is, were favorable toward programmed instruction. An analysis of variance showed that significant differences occurred between types of consumers but that any differences between schools were not significant.

Another question asks:

"Did you miss not being able to participate in class discussions or miss not asking questions as you would have in a conventional class?"

An analysis of variance showed that differences exist between types of consumers but not between schools. The individual group means for most instructor and administrator groups fell below the neutral point while the means for most student groups were above the neutral point. Apparently, the students do not miss class discussion as much as the instructors and administrators think they do.

Another question reads:

"How do you rate programmed instruction for learning new material?"

The analysis of variance showed that all differences were significant, between types of consumers, schools, and interactions. All groups scored above the neutral point except the administrators at the Photographer's School. Additionally the students were consistently more favorable toward programmed instruction than instructors or administrators except for the Aviation Ordnance School where the students and instructors had equal means.

A companion question to this last one asks: "How do you rate programs for reviewing materials?"

All groups scored above the neutral point for this question. The analysis of variance showed that the only differences that were significant were those between schools.

As for reviewing material from the programs, at the end of each booklet that we use, there is a test which the student takes after completing the program. This helps the student determine the areas in which he is weak. In addition, about half of these tests provide the students with page numbers to review for each incorrect answer. These tests are considered very helpful and may contribute to the extremely favorable responses on this question by the instructors and administrators.

The next question asks: "To what extent do you find the programs repetitious?

1. Not at all repetitious.
2. Slightly repetitious.
3. Uncertain.
4. Too repetitious.
5. Much too repetitious."

A companion item states: "Programmed instruction is a boring method of learning." The subjects were asked to respond:

1. Strongly agree
2. Agree
3. Uncertain
4. Disagree
5. Strongly disagree

Analyses of variance showed significant differences occurred between schools for both items. Two student groups fell below the neutral point for the "repetitious" question and three student groups fell below the neutral point for the "boring" question. These results may be interpreted as showing that the quality of the programs in the schools vary to some extent and may point to a need for improving this aspect of the programs at some of the schools.

The next question asks about retention of programmed materials compared to retention of conventional materials.

Each of the student groups scored above the neutral point, that is, were more favorable toward programmed instruction. However, the instructors and administrators at over half of the schools felt that programmed instruction was inferior to conventional instruction in this regard. An analysis of variance showed that differences in attitudes did exist between types of consumers and between schools. It is interesting to note here a study by Dr. Kirk Johnson which showed that students did as well or better on a final examination after taking a course which included programmed instruction than after taking the same course using only conventional instruction.

Another statement reads: "Programmed instruction is bad because there is never a teacher around to explain anything." The subjects are again asked to respond:

1. Strongly agree
2. Agree
3. Uncertain
4. Disagree
5. Strongly disagree

An analysis of variance showed that significant differences occurred between types of consumers on this item. From the individual means, it appears that the instructors feel that they are available to the student more often than the students feel they are. This is an area that will have to be checked more closely.

Finally, the next item was used as a summary question. It asks: "In terms of an overall evaluation of programmed instruction at this training activity, I believe that programmed instruction:

1. Is making (or will make) a major contribution.
2. Is making (or will make) a contribution of some importance.
3. Is no better or worse than what we have had before.
4. Is (or will be) somewhat detrimental.
5. Is (or will be) very detrimental."

Although the analysis of variance showed that differences between types of consumers and interaction differences did occur, the individual group means for all groups fell above the neutral point. It is clear that all consumers felt that programmed instruction is making a worthwhile contribution to the training program in the Naval Air Technical Training Command.

SCHOOL OF CONTINUING EDUCATION RESEARCH PUBLICATIONS
(Formerly University College Research Publications)

1. Strengthening the Residential Adult Education Experience. Sponsored by the Conference and Institute Division of the National University Extension Association, The Center for the Study of Liberal Education for Adults, and University College, November 1962, 62 pp.

2. Objectives and Methods of Research in Adult Education. Edited by Philip H. DuBois and King M. Wientge, May 1962, 51 pp.

3. Studies of Biographical Data in Adult Education. Philip H. DuBois and King M. Wientge, August 1963, 8 pp.

4. Strategies of Research on Learning in Educational Settings. Edited by Philip H. DuBois and King M. Wientge, February 1964, 74 pp.

5. Factors Associated with the Achievement of Adult Students. King M. Wientge and Philip H. DuBois, 1964, 39 pp.

6. Survey of Tuition Aid Plans of Business, Industry, and Government in Metropolitan St. Louis. King M. Wientge and Malcolm Van Deursen, June 1965, 12 pp.

7. Workshop for Counselors and Educators Concerned with the Education, Training and Employment of Minority Youth, Final Report, Part I: Development, Program, Evaluation. King M. Wientge, October 1965, 71 pp.

8. Workshop for Counselors and Educators Concerned with the Education, Training and Employment of Minority Youth, Final Report, Part II: Discussion Guide to the Problems of the Culturally Deprived: An Introduction for Teachers and Counselors. John M. Whiteley and King M. Wientge, November 1965, 71 pp.

9. Criteria in Learning Research. Edited by King M. Wientge and Philip H. DuBois, 1966, 65 pp.

10. The Influence of Social Climate on Adult Achievement: The Impact of a Residential Experience on Learning and Attitude Change of Adult Students Enrolled in an Evening Credit Class. King M. Wientge and James K. Lahr, 1966, 26 pp.